

Collaboration *in extremis*

Conditions could not be worse for collaborative science between Israel and its neighbours. But the interests of politics, regional development and science all favour such collaborations. Institutions as well as individuals should promote them.

The Israeli peace movement, relegated to silence following the escalation of conflict since the second intifada in September 2000, reasserted itself last weekend when almost 100,000 people turned up for a pro-peace demonstration in Rabin Square in Tel Aviv. Support for a just peace by progressive academics in the region through scientific collaboration has suffered. They now need the active support of the international science community to resuscitate and sustain collaboration. At stake is not only the contribution that this can bring to the peace process, but also the genesis of a viable scientific and technological base in Palestine and other Arab countries, which could play a key role in ensuring both future prosperity and a joint approach to common regional problems.

One impact of the profound polarization of Israeli and Palestinian societies is that, whereas collaborating scientists were once proud of their activities — often publicized by their institutions in the hope of attracting international funding — many now adopt low profiles. But collaboration between the research powerhouse that is Israel, which normally has a tendency to look to the West for cooperation, and its Arab neighbours needs to be protected as much as possible from the impact of the political conflict.

Science is one of the few areas in which Israelis and Palestinians can interact apolitically, and the importance of the trust and dialogue created by such second-track diplomacy to the political peace process should not be underestimated. Many Israeli scientists are committed to this cause and, more self-interestedly, collaboration is a means for Israel to win something it longs for: acceptance by Arab states. Palestinians, on the other hand, recognize that collaboration with Israel is key to gaining access to the scientific facilities they lack and the training of a new generation.

Track records

And there is the science itself. Almost two-thirds of joint projects involve applied science relevant to critical regional problems — for example, in medicine, agriculture, water and the environment. The Kuvin Center for Infectious and Tropical Diseases of the Hebrew University of Jerusalem and Hadassah Medical Organization has long collaborated with Arab scientists on diseases such as malaria and leishmaniasis. In one of its most notable successes, researchers at Ain Shams University in Egypt, with funding from the US Agency for International Development and the US National Institutes of Health, identified and developed a vaccine for Rift Valley fever, a Mediterranean disease that is fatal to humans and livestock. The collaboration published over fifty joint papers.

For all these reasons, international scientific communities need to redouble their efforts to promote peace and science through greater collaboration. What can be done? Benchmarking of what has been achieved so far, what the problems are, and how funds can best be invested is an essential and neglected task. Funds could be used to train Palestinian scientists and allow them to attend meetings and workshops. A high-level 'Science for Peace' conference, bringing together the main players from both sides, would also be worthwhile. More researchers outside the region could initiate three-way collaborations. Now is a good time to contact

scientists in the region to offer support and learn of their needs.

Such ambitions are controversial (see page 221) and cannot ignore the painful practical realities of politics and security, not to mention the devastation of Palestinian infrastructure (see page 209). Funders should bring whatever influence they can bring to bear to minimize such damage and improve conditions.

Restrictions on freedom of movement are viewed by Palestinian researchers as the main obstacle to greater collaboration with Israel. Even in more peaceful times, Israel's arbitrary approach to awarding Palestinians permits to travel within Israel has been a source of humiliation and wasted effort for Palestinians. Israeli researchers may travel anywhere within Palestine, but Palestinian researchers wishing to visit their colleagues in Israel must apply for special permits — and, they say, depending on the mood of the Israeli official on duty, it may be for one month or one day. And they cannot stay in Israel after dusk.

Mutual respect

One Palestinian scientist sums up the resentment: "We want to cooperate but why should we go and normalize our relationships when the Israelis here make our life miserable, when we can't cross the border to meet in Hebrew University, which is just down the road?" To their credit, individual Israeli scientists often go out of their way to help them obtain permits. Security concerns will be paramount, but outside pressure needs to be applied to encourage greater action by Israeli research organizations, who have failed to use their considerable political clout to obtain decent travel terms for their Palestinian colleagues.

Similar concerns arise within science itself. When Israelis speak of collaboration, many Arab scientists fear that this will be overly dominated by Israelis' superior research infrastructure and funds, with Arabs used as technicians. What Palestinians want is better access to that rich infrastructure, and true collaboration based on 'mutual respect' and 'dignity'. Such terms are often vacuous, but in the context of the Middle East, they represent a profound aspiration that should not be ignored.

The term 'normalization' is key to understanding Israeli-Arab scientific relations. Palestinian universities refuse formal institutional links with their Israeli counterparts on the grounds that these are appropriate only after significant progress has been made towards peace. Many Israelis likewise prefer personal collaborations. But ultimately there needs to be a commitment to formal collaboration between institutions. Only then can the full potential of research collaboration be realized. Is it too much also to hope that institutions might take a public stand on issues of human rights and equality that are hindering greater Arab-Israeli collaboration and dialogue?

Palestinian universities such as Birzeit have struggled admirably to educate the next generation of qualified Palestinians, despite having often been cut off from the outside world for months at a time. The silence of its Israeli counterparts, who share the same academic values, in reaction to its plight says much about how far the conflict has damaged the values on which the scientific and educational enterprise depends. ■





Screen test
Standards set
for assessing
mutant mice
p211



Gun law
Utah university
stands firm over
weapons ban
p212



Jailhouse mock
TV repeat of prison
experiment sees
guards go soft
p213



Love birds?
Feathers fail to fly as
world's rarest birds
choose celibacy
p214

Science collaboration stymied by relentless Middle East conflict

Declan Butler, Paris

Years of painstaking effort to boost scientific collaboration in the Middle East have been battered — but not yet destroyed — by the latest conflict in the region, researchers say.

War conditions in the region, supply shortages and restrictions on travel by Palestinian scientists have paralysed collaboration, and the escalation of conflict has polarized the views of researchers in both camps.

Nonetheless, many researchers are convinced that any let-up in the conflict may allow for a rapid revival in collaboration between Israeli and Arab scholars. There are also signs that the current crisis will galvanize international interest in promoting such collaboration.

In the optimism that followed the signing of the 1993 Oslo peace agreement, science emerged as one of the few areas where Arabs and Israelis could cooperate on common goals in a non-political setting. In this way, it could serve as a bridge for dialogue and mutual understanding (see *Nature* **375**, 717–732; 1995).

But now, with the peace process in tatters, collaboration has been hit hard. Official scientific organizations in the region are under



Shattered hopes: Israeli and Palestinian scientists must work to repair damage done by the conflict.

attack for their conspicuous public silence on the issues that matter most to researchers on both sides. Many observers argue that these organizations need to take a more courageous public stand if impetus is to be regained.

Palestinian scientists cite travel restric-

tions imposed by Israel as a major impediment to collaboration. These have become steadily stricter since the start of the second Palestinian uprising, or intifada, in September 2000. And since the occupation of the West Bank, movement for Palestinian researchers and students there has become almost impossible.

"The upper levels of Israeli universities were pretty much unwilling to use their political clout in pushing for freedom of movement," says Paul Scham, a researcher at the Harry S. Truman Research Institute for the Advancement of Peace at the Hebrew University of Jerusalem and the author of a study on Israeli/Arab research cooperation.

"Israeli scientists tend to distinguish between support for academic cooperation and ending frontier closures, which they consider a security matter. But Palestinians consider greater freedom of movement a prerequisite to more normal relations with Israeli academics," says Scham.

Israeli researchers, meanwhile, complain privately that their Palestinian colleagues have said nothing in public about the suicide bombings that provoked the occupation of the West Bank.

As Israeli tanks withdrew from the West Bank earlier this month, tensions between

Plan hatched to double NSF budget

Geoff Brumfiel, Washington

The Science Committee of the US House of Representatives is leading a new charge to double the \$5-billion annual budget of the National Science Foundation (NSF) within five years.

Scientific societies are strongly backing a bill, passed last week by the committee with bipartisan support, that would recommend more money for the NSF, the largest sponsor of non-biomedical university research in the United States. A companion measure to the bill is being planned in the Senate.

The NSF has won plaudits from the Bush administration for its management skills, but this year's budget includes only a modest 5% funding increase (see *Nature*

415, 565; 2002). "Recognition is nice, but success requires real money," says Sherwood Boehlert (Republican, New York), chair of the Science Committee.

If it is passed into law, the bill would recommend 15% increases in the NSF budget for each of the next five years. Actual increases are set each year by appropriations committees, which are likely to be influenced by such a law.

Congress has made statements supporting the doubling of funding for research, but only the National Institutes of Health, which supports biomedical research, has actually obtained the money. Now science lobbyists are hoping that the NSF can garner similar treatment. ■

the Palestinian and Israeli science communities heightened further. Palestinian officials and Israeli peace activists claimed that damage done during the occupation had weakened the fledgling Palestinian research and education infrastructure.

The Palestine Academy of Science and Technology in Ramallah, for example, was damaged when Israeli troops used explosives to blast open doors before searching the building, says Imad Khatib, director general of the academy. Furniture was smashed, and files and office equipment were scattered on the floor, he adds.

Al Quds University claims that Israeli soldiers badly damaged laboratories and other buildings at its campuses in El Bireh and Ramallah. The university has asked the Israeli government and the international community to send fact-finding missions and to help rebuild its infrastructure. It has also pleaded for US\$675,000 of emergency aid, without which it says it will be unable to pay its 700 staff and may face closure.

Israeli officials were unavailable to comment on the specific allegations as *Nature* went to press. But the Israeli army has said it is taking allegations of vandalism seriously and will investigate them fully. The extent of the damage and the circumstances in which it was inflicted remain unclear. "It has been difficult to isolate and concentrate on specific aspects of the military operation," says Eva Illouz, a peace activist and researcher in sociology and anthropology at the Hebrew University. "News of the destruction of the research facilities went largely unnoticed."

Palestinian scientific leaders accuse the Israeli scientific community of being oblivious to their plight. Khatib, for example, complains that Israeli scientific leaders are "indifferent" to the damage at the Palestine Academy. Illouz counters that Israeli scientists are "disturbed" by reports of the destruction, but admits that there has been no public outcry because attitudes have hardened between Arab and Israeli scientists following the second intifada.

Dudy Tzfati, a geneticist at the Hebrew University of Jerusalem and one of the founders of the Alliance of Middle Eastern Scientists and Physicians, a body that is committed to promoting peace, nonetheless remains optimistic about the future. He predicts that the mutual trust created by contacts among scientists will endure, and that the international scientific community should renew its support for such cooperation.

Similarly, Nachum Finger, rector of Ben Gurion University, predicts that, on any sign of a return to normality, "the universities, the scholars, will be back to lead the *rapprochement* with the Palestinians". ■

Warhead-saving arms pact gets lukewarm reception

Geoff Brumfiel, Washington

American scientists have reacted with little enthusiasm to a new US–Russian nuclear arms pact, announced on 13 May, that would reduce the number of nuclear weapons deployed on each side but would take no action to destroy their warheads.

The three-page treaty, which will be signed by presidents George Bush and Vladimir Putin at a summit in Moscow next week, pledges to reduce the number of deployed nuclear weapons from 5,000–6,000 on each side to between 1,700 and 2,200 over the next 10 years.

But the plan would allow the decommissioned warheads to be kept in reserve rather than being destroyed — raising alarm about the possible security of the warheads, particularly in Russia. The freedom to keep the warheads was sought by the Bush administration, but it runs counter to the advice of many US scientists. In a 1997 study, the National Academy of Sciences called for the bilateral destruction of decommissioned warheads.

Scientists welcomed the fact that the two heads of state will actually sign a treaty — the Bush administration had originally proposed a less formal agreement — but lamented its

lack of real content. On the day the treaty was announced, the Union of Concerned Scientists released a letter signed by seven prominent US physicists to both presidents, calling for a treaty that would destroy decommissioned warheads as quickly as possible.

"This treaty is certainly a step forward, but it leaves a lot to be desired," says Michael Levi, a physicist at the Washington-based Federation of American Scientists. Levi says that the agreement will strengthen pro-Western elements in the Russian government. But he adds that, in storage, Russian warheads pose a significant threat. "If one of those warheads falls into terrorist hands, then we'll wish that we had used our stored warheads as a bargaining chip to have them eliminated," he warns.

Wolfgang Panofsky, former director of the Stanford Linear Accelerator Center in California, and one of the seven who signed the letter, says he is disappointed that the treaty fails to address smaller tactical warheads, which, he says, are considered a significant threat to non-proliferation efforts. Nonetheless, he says, "the treaty should be applauded as a forward step". ■

♦ www.ucsusa.org

Computer revives Bamiyan Buddha

Quirin Schiermeier, Munich

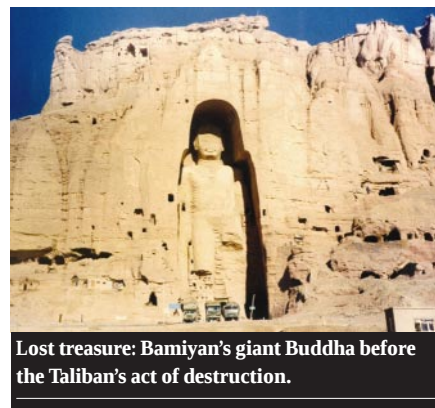
The giant statue of Buddha at Bamiyan in Afghanistan, one of two destroyed in March last year by the Taliban, will soon be reconstructed — electronically.

Researchers at the Swiss Federal Institute of Technology in Zurich are close to completing a three-dimensional computer model of the 2,000-year-old statue, using a set of high-resolution images taken in 1970 by Austrian cartographers.

"A high-accuracy computer model, which could be used as the basis for physical reconstruction, will be available by the end of the month," says Fabio Remondino, one of the scientists involved in the project.

But authorities warn that a full physical reconstruction is unlikely. Christian Manhart, an Asian specialist at the United Nations Educational, Scientific and Cultural Organization, says that rebuilding the statue, which stood nearly 55 metres tall, is neither realistic nor desirable.

The site should be left as it is and be protected as a memorial, similar to the concentration camp at Auschwitz or the memorial church in Berlin, Manhart says, as



Lost treasure: Bamiyan's giant Buddha before the Taliban's act of destruction.

recommended by the 1964 International Charter for the Conservation and Restoration of Monuments and Sites.

Paul Bucherer-Dietschi, director of the Afghanistan Institute and Museum near Basel in Switzerland, is coordinating a feasibility study on possible reconstruction of the statue. He will meet with 40 other experts in Kabul this month to discuss the protection and restoration of Afghanistan's cultural heritage. ■

♦ www.photogrammetry.ethz.ch/research/bamiyan

Geneticists lay plans for rationalized rodents

Alison Abbott, Munich

Mouse geneticists in Europe have launched an initiative to standardize the flood of data being generated by studies of mouse mutants.

At a meeting in Paris last week, the heads of European mouse-genetics research centres agreed to unify the screening processes used to identify mouse mutants that may aid studies of human diseases.

The 17 members of EUMORPHIA — the European Union Mouse Research for Public Health and Industrial Applications network — plan to use two levels of screens, and to make full descriptions of the mutants available on a public database.

"The first level would hopefully be applied to all mutants," explains Steve Brown, director of the Medical Research Council's Mammalian Genetics Unit at Harwell, UK, and one of the two researchers running EUMORPHIA. "We'll design a fast series of tests to answer simple questions. Are the mice deaf? Are they overweight?" Second-level tests will then be used to take a more thor-



Screen star: standardized tests for mutant mice may speed up research into human diseases.

ough look at the unusual characteristics identified in the first screen.

In the past four years, hundreds of mouse mutants have been created by random chemical mutagenesis, in which chemicals are used to introduce a random mutation into the animal's genome. Researchers study the effect of the mutation and, if the resulting characteristics are interesting, identify the gene responsible.

Knockout mice, in which a specific gene is targeted and disabled, are also widely used to model and study human diseases. The rate at which both types of mutant are being created has been accelerating as the mouse genome sequence nears completion.

Two centres — the GSF national research centre in Munich and the Harwell unit — have been screening mutants generated by chemical mutagenesis, and have already identified nearly 600 types that may be useful in disease studies.

Some of the standardized tests will be based on those already used at Munich and Harwell, with the task of defining the others being split between EUMORPHIA members. Experts from outside Europe will also be invited to participate, and EUMORPHIA members hope that the tests will soon be adopted globally.

Knockout mice could also be screened using the new tests, says Pierre Chambon, director of the Institute of Genetics and Molecular and Cellular Biology in Strasbourg, France, and the other EUMORPHIA leader.

This could be beneficial, he adds, recalling an incident when he discovered by chance a link between retinoic acid and the neurotransmitter dopamine. While studying developmental problems, one of the knockout mutants he produced showed some symptoms reminiscent of Parkinson's disease — in which the brain's dopamine system is defective. "It turned out that one of the brain's dopamine receptors is regulated by an element controlled by retinoic acid," says Chambon. He hopes that such chance findings will become routine if knockouts are screened. ■

Unrest leads reformer to quit Seoul university's top job

David Cyranoski, Tokyo

The president of Korea's largest university has resigned in the face of student unrest and faculty opposition — throwing open to question his plans to turn Seoul National University into an international research powerhouse.

Students and some faculty members had accused Ki-jun Lee of running the university autocratically while maintaining closer links to business than is permitted in Korea. But many scientists and engineers believe he was ousted because his efforts at reform upset some staff, especially in the university's humanities departments.

Instead of completing his four-year term this November, Lee resigned on 1 May and will return to his research in the university's chemical-engineering department.

Student protests against Lee's plans to raise tuition fees and implement other reforms included a 10-day occupation of his office last month. In addition, students and representatives of two faculty associations alleged that Lee had used university funds for private purposes, although his staff say

that his expenditures were in line with university policy.

Critics also attacked him for retaining a paid directorship at LG Chemical, a petrochemicals firm. Korean law forbids professors at public universities from accepting payment for work in industry. An official on his staff says Lee did not violate any law, as the money from the directorship

was to support company-related research.

Some university researchers and industry officials say that the main source of opposition to Lee was anxiety among arts professors and some scientists about his reforms. "Professors of basic natural science, humanities and social sciences just didn't like his emphasis on practical sciences and engineering," says one university researcher.

Lee's supporters say that his reforms were working. They claim, for example, that the university has risen substantially in international rankings that measure the number of times its scientific papers have been cited in quality journals.

In Korea, where university presidents usually lack strong executive powers, some fear that the resignation sets a grim precedent for Lee's successor. "The next president will have a difficult time," says the researcher.

"He wanted to remake the tenure system," says the president of a major Korean company who declined to be identified. "He was trying to exercise some muscle, which university presidents don't normally do." ■



Utah law triggers university row over concealed weapons



UNIV. UTAH; CORBIS (INSET)

Rex Dalton, San Diego

Parking for the staff and sports for the alumni may be the hottest issues troubling some university presidents. But Bernard Machen, president of the University of Utah, has a more pressing concern: a push by the state government to make sure that everyone on campus feels free to carry a concealed firearm.

Scientists at the Salt Lake City campus are up in arms themselves over the state's push to ensure that guns are permitted, and seen to be permitted. "The situation is bizarre," says Rob MacLeod, a bioengineer at the university who specializes in electrocardiology. "I tell my friends about this and they look at me like I am from another planet."

The drive is backed by the state legislature and the Utah attorney general, Mark Shurtleff. "The fear of concealed weapons is groundless and irrational," Shurtleff says. But some professors say they will consider resigning or retiring if the state is successful.

The issue has arrived on campus through a wider movement by US gun advocates and conservatives to enact state laws permitting the carrying of concealed weapons as a potential deterrent to crime. Like many states, Utah passed such laws several years ago, allowing anyone who is over 21 years old and not a felon to obtain such a permit. The permits are easily acquired by most adults.

Shurtleff issued an opinion last November saying that nine state-funded colleges cannot have policies that prohibit the carrying of concealed guns. Utah's effort to enforce the concealed-weapon law on campuses is thought to be the first of its type.

The University of Utah — the largest and most powerful of the nine — is trying to take a stand against the Shurtleff ruling. In March, it asked the US District Court in Salt Lake City to issue an opinion to allow it to enforce a policy banning concealed weapons on campus. No hearing has been set yet, but the university is hoping for a decision late this year.

None of the other eight state-funded



Fire fighting: the University of Utah (top) wants to stop students from carrying guns on campus.

universities has joined the court action — perhaps fearing financial retribution from the legislature, which holds their purse strings — and all of them are permitting concealed weapons on campus while they await the court's decision. Earlier this year, a committee of the legislature narrowly rejected a motion to cut Machen's pay in retaliation for his stance on the issue.

In documents submitted to the court, the university argues that it should be allowed to limit concealed weapons in order to create an environment in which people can exchange ideas freely, without fear of being shot.

"I don't see handguns being appropriate for classroom discussions," says computer scientist Christopher Johnson, director of the Scientific Computing and Imaging Institute at the university.

Professors and others opposed to the proposal are trying to collect the 30,000 signatures needed to force a statewide vote this November on a resolution that would exempt campuses from Utah's concealed-weapons law.

"We're concerned about a tragedy in the making here," says chemist Peter Stang, dean of the university's College of Science. Johnson says that his wife, Katharine Coles, an English professor who is president-elect of the University of Utah's faculty senate, has been told that "distinguished members of the faculty have indicated they will leave" if the university is not successful in banning concealed weapons.

And ecologist Phyllis Coley says: "A colleague from New York called to say he is fearful of coming here to give a talk." ■

Deep-sea nuggets prove irresistible to Indian miners

K. S. Jayaraman, New Delhi

India is set to go ahead with plans to mine millions of tonnes of manganese nodules from the Indian Ocean's floor, despite warnings from an environmental assessment that the project will disrupt deep-sea ecology.

The potato-shaped nodules are rich in manganese, copper, nickel, cobalt and iron. India has a shortage of some of these elements, and has purchased exclusive rights from the International Seabed Authority to mine the nodules from 150,000 square kilometres of the Indian Ocean. The country estimates that the reserves could be worth as much as \$200 billion.

But researchers at the National Institute of Oceanography (NIO) in Goa, who undertook a survey to assess the environmental impact of mining operations, say that discharges of unwanted sediment during mining will upset the area's ecological balance.

To simulate the mining process, the researchers mimicked the release of sediment over a small area 5,500 metres below the sea's surface.

During a nine-day experiment in 1997, they lifted about 580 tonnes of sediment from the sea floor and then released it five metres above the sea bed. They have since been tracking how the sediment plumes drift and what effect they have on the area's ecology.

The recently released report says that "noticeable changes were observed in the physical and biogeochemical properties" of the area. The sediment plumes migrated at least two kilometres from the drop site, and the abundance of fauna and bacteria decreased in the region, says marine geologist Anil Valsangkar, who led the study.

Valsangkar says that actual mining would cause much more disruption if, as is likely, the mining ships discharged the sediments at the sea's surface.

The Indian government will not say when it plans to begin the mining operation, but it has spent about \$40 million surveying the site, has bought a ship and is developing machinery for the purpose. NIO sources say that the environmental study was conducted to comply with the law, rather than to obtain clearance to proceed.

But critics have questioned the plan's economics, given that the nodules are so far below the sea's surface and that the site is 2,000 km away from India. ■

Reality TV show recreates famed social study

David Adam, London

Final proof that scientists can work with the media, or an unfortunate clash of peer pressure and peer review? Television viewers in Britain can make up their own minds this week when a unique combination of science and reality TV gets its first screening.

Billed as both entertainment and serious research, the BBC programme *The Experiment* attempts to recreate aspects of a 1971 psychology study by Stanford University researcher Philip Zimbardo. In the original study, volunteers were asked to assume the role of either prisoner or prison guard, and were placed together in a simulated jail. Those acting as guards were so brutal to the prisoners that a shocked Zimbardo abandoned the experiment and said that it should never be repeated. It is now seen as an important demonstration of how social circumstances can make people conform to particular roles.

But whereas the Stanford guards bullied the prisoners and forced them to clean toilets with their bare hands, their British counterparts tried to befriend the prisoners and even agreed to reorganize the prison as a commune. An edited version of the 10-day trial, held last December at a London television studio, will be shown in four one-hour programmes beginning this week.

"This has been a ground-breaking collaboration between academia and the media," says Alex Haslam, a psychologist at Exeter University who planned the experiment with Steve Reicher, a researcher at the University of St Andrews and joint editor of the *British Journal of Social Psychology*. "It is not a media creation on which psychologists were invited to comment. The guiding principle for us and the BBC was always the science."

Others warn, however, that it is difficult to compare the two studies directly, as behavioural trials are hard to replicate. The television cameras will also have had "an immediate constraint on what people do", says Donald Laming, a psychologist at the University of Cambridge. Zimbardo, still at Stanford University, could not be reached for comment.

For *The Experiment*, 15 male volunteers were recruited through newspaper adverts, and divided into 6 guards and 9 prisoners. The prisoners shared cramped cells and carried out chores, whereas the guards enjoyed better food and separate rooms and were given limited powers over their charges. As well as observing and recording the behaviour of the volunteers, Haslam and Reicher asked them to complete psychometric tests and analysed their saliva for the stress hormone cortisol.

It soon became clear that the guards were



Broadcasting behind bars: *The Experiment* aims to explore how people conform to set roles.

unable to organize themselves. "There was a lack of a common social identity, no leadership and they didn't even develop a simple shift pattern," says Reicher. The guards showed more psychological stress than the prisoners.

Haslam says that he welcomed the chance to challenge the Stanford results, and claims that the guards who behaved as tyrants in the original experiment did so only because they were encouraged to.

BBC

Jefferson's descendants continue to deny slave link

Erika Check, Washington

Thomas Jefferson's descendants have decided that the ancestors of one of his slaves cannot join their family club, despite DNA evidence published in *Nature* suggesting that he fathered at least one of her children.

The Monticello Association, a group of nearly 800 descendants of Jefferson's daughters, voted on 5 May to keep Sally Hemings's descendants out of their group. This means that Hemings's relatives cannot

be buried in the graveyard at Monticello, Jefferson's home in Charlottesville, Virginia.

Their vote is the latest twist in a racially charged controversy that has raged since Jefferson's own day, when it was noted that Sally Hemings's children bore a likeness to the third president. That and other evidence fuelled speculation that the author of the Declaration of Independence fathered six children by Hemings, who was a slave at Monticello. She only became pregnant during times when Jefferson stayed there.

Jefferson's descendants have always disputed this speculation. Jefferson's grandchildren claimed that one of his maternal nephews, Samuel or Peter Carr, fathered Hemings's children. But genetic tests published in November 1998 (see *Nature* 396, 27–28; 1998) eliminated that possibility. The tests compared the Y chromosomes of men descended from Sally Hemings's son Eston with those from the Jefferson and Carr family lines. Eston's Y chromosome was found to be virtually identical to Jefferson's and dissimilar to the Carr chromosome.

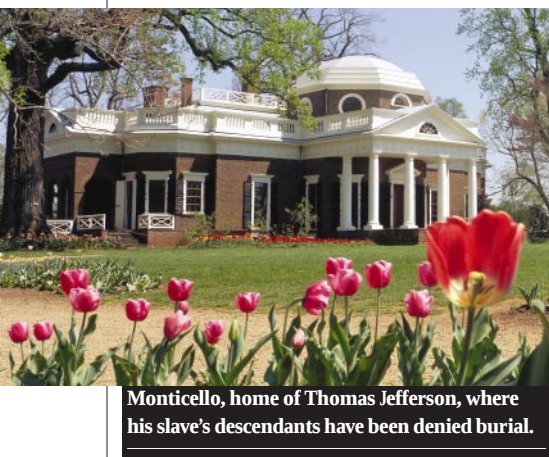
The study caused an outcry because it was published in the midst of then-President Bill Clinton's impeachment process. Clinton's

supporters seized on it as proof that previous presidents — even the most celebrated ones — had also engaged in sexual indiscretion.

The findings prompted the Thomas Jefferson Foundation, which owns and operates Monticello, to investigate the Hemings connection. In January 2000, it concluded that the DNA study, combined with other evidence, indicated a "high probability that Thomas Jefferson fathered Eston Hemings, and that he most likely was the father of all six of Sally Hemings's children".

But Jefferson's descendants remain unswayed, and now say that Thomas's brother Randolph was probably Eston's father. "They have not provided conclusive evidence that Thomas Jefferson had a relationship with Sally Hemings," says Nathaniel Abeles, the Monticello Association's president.

The association's position cuts no ice with the academic community. "This decision is scientifically unsound, historically unsound and morally bankrupt," says Annette Gordon-Reed of the New York Law School. "The issue is pretty much settled and the fact that the family association chooses to tell a story about itself doesn't affect what scientists or historians think."



Monticello, home of Thomas Jefferson, where his slave's descendants have been denied burial.

J. SOHIM/CHROMOSOM/CORBIS

White House sets up panel to vet foreign student visas

Washington Plans to vet foreign students applying to study at US universities were announced last week by the White House.

A new committee, the Interagency Panel on Advanced Science and Security, will review applications for courses such as nuclear physics and engineering, which contain content that could be useful to terrorists.

Government officials say that the panel, which will come under the control of the State Department and the Immigration and Naturalization Service, will examine a small number of applications to determine whether a visa should be issued. Applications will be chosen for review on the basis of the student's personal background and country of origin.

News that the US government was planing to vet applications from foreign students had led to fears that the new controls could stifle scientific exchange (see *Nature* **416**, 111; 2002). But many scientific associations were reportedly satisfied with the case-by-case approach to be taken by the new committee.

Selective genetics exhibition leaves out Mendel's religion

Munich Science has triumphed over religion in an argument over an exhibition about Gregor Mendel, the Augustine monk who discovered the laws of heredity.

Mendel's former home, the Abbey of St Thomas in Brno in the Czech Republic, will host the new exhibition, which is due to open next week. The current abbot had asked that the religious background to Mendel's life and work form part of the exhibition (see *Nature* **410**, 6; 2001), but dropped this request after the abbey agreed to host annual workshops on bioethics.

Kim Nasmyth, scientific director of the Vienna-based Research Institute of Molecular Pathology, who helps to promote genome research in Brno, says that the temporary exhibition will eventually become a permanent museum of genetics, possibly with affiliated laboratories.

Movie mogul's millions make UCLA's day

San Diego Hollywood executive David Geffen once used a fake degree from the University of California, Los Angeles, to land his first job in the entertainment industry. Now he has repaid his fictitious *alma mater* with a donation of \$200 million.

Geffen, who is now a partner in the DreamWorks SKG entertainment company



Geffen in 1973: the 'degree' was fake but his talent proved real.

mailroom of a major talent agency. He went on to become a billionaire after founding and operating music, movie and theatre ventures. He has also made contributions to AIDS research.

I think, therefore I am head of French research (for now)

Paris For the next month at least, French research has a philosopher at its head and an engineer at the controls.

Luc Ferry, a renowned specialist in the philosophy of politics, was named minister of a new superministry that brings together education, youth and research. The ministry forms part of the new interim government formed by Prime Minister Jean-Pierre Raffarin after Jacques Chirac's re-election to the French presidency on 5 May. Ferry is supported by François Loos, an industrial engineer who has been given responsibility for research and higher education within the new ministry.

Ferry and Loos, both of whom have advised previous governments, may only have five weeks to reflect on the future of French research and education. Next month's legislative elections will install a government for the coming five years, and could result in a change in ministers.

Cash cut hits Italian neutrino detector

Rome Physicists in Italy have been forced to rethink a major neutrino experiment after CERN, the European particle-physics laboratory, cancelled its support in response to financial difficulties.

The OPERA experiment is designed to study a beam of neutrinos sent from CERN in Geneva, Switzerland, through the Alps to the Gran Sasso laboratory in central Italy. CERN agreed in 1999 to help fund the experiment (see *Nature* **402**, 847; 1999), which is intended to probe oscillations between different varieties of neutrino. The laboratory now says that it is withdrawing

of Los Angeles, announced he was making the donation on 7 May. The money, thought to be the largest ever donation to a US medical school, will be used for research and teaching at the renamed David Geffen School of Medicine.

Biographies of Geffen record that in 1964 he claimed that he had a degree in theatre arts in order to win a job in the

support for the detector in order to save SFr12.5 million (US\$7.8 million), but that it will continue to build the beamline.

CERN is currently reviewing its financial commitments following the announcement earlier this year that its next major project, the Large Hadron Collider, was some SFr850 million over budget (see *Nature* **413**, 441; 2001). Paolo Strolin, a high-energy physicist at the University of Naples and spokesperson for the OPERA experiment, says that the decision was not unexpected, and that OPERA researchers are now trying to simplify the detector's design without compromising its performance.

No love nest for rarest birds as they fly towards oblivion

Maui An attempt to bring together a pair of the world's rarest known birds has failed after the female headed home without meeting her intended mate.

Only three po'ouli (*Melamprosops phaeosoma*) are thought to survive on the Hawaiian island of Maui. The home ranges of the two surviving females and one male do not overlap, stunting the love lives of the finch-like birds. Earlier this month, researchers caught one of the females and released her in the male's home range at dusk. But she was on her way home by the middle of the next day, apparently without having encountered the male.

"It's obviously disappointing," says Jim Groombridge, head of the Maui Forest Bird Recovery Project. With the breeding season over, he and colleagues are considering their next move — probably captive breeding, either away from the forest or in an on-site aviary. But the birds are getting old. "We do not have a great deal of time left," he says.

Birds have come back from the brink of extinction before. During the 1970s, the population of Mauritius kestrels fell to four, and that of New Zealand's Chatham robins to six. Thanks to captive-breeding programmes, there are now hundreds of each species.



Safe hands: but are human efforts enough to help the last of the po'ouli ward off extinction?

AP

C. BROSTIUS

What kind of science is this?

Mathematical prodigy Stephen Wolfram has laboured for a decade on what he claims is a revolutionary book. Jim Giles meets a supremely confident scientific loner, but finds expert opinion on the work's merits divided.

To all intents and purposes, Stephen Wolfram dropped out of the research community more than a decade ago. Since 1988, he has neither published a scientific paper nor attended a conference. But the British-born mathematician has not been idle. Working long into the night, gazing into the screen of his computer at his Chicago home, Wolfram claims to have sown the seeds of a scientific revolution.

The fruits of this solitary labour are revealed this week in the mammoth tome *A New Kind of Science* (Wolfram Media, Champaign, Illinois, 2002). It is a call for researchers to turn away from calculus and other conventional mathematical tools and to embrace instead simple 'rules' that can be applied to generate patterns of astounding variety and complexity. Hidden within these patterns, Wolfram asserts, are the keys to understanding a multitude of biological and physical phenomena from the shapes of leaves to the structure of space-time itself. He suggests that his work will change almost every branch of the natural sciences, and even social sciences and the arts. "All the media are going to follow this — and in a big way," he predicts.

Coming from most authors, such grand claims would instantly be dismissed as empty hype. But Wolfram's track record will ensure that many scientists will reserve their judgement. A bona fide prodigy, Wolfram published his first paper on theoretical high-energy physics aged 15, and later breezed through a PhD in a year. Before turning 30, he had helped to launch the discipline of complex-systems research and had founded a mathematical software company that has



D. REISS/WOLFRAM RESEARCH

since made him — at the still relatively tender age of 42 — a very wealthy man.

The book's title typifies a brash approach that many characterize as arrogance. "I was successful in science early in life," Wolfram says matter-of-factly, over a restaurant meal in Cambridge, Massachusetts. "It leads to a certain degree of self-confidence." But do the book's 1,200-odd pages really detail a new way of doing science? Or has Wolfram's supreme self-belief, unfettered by the need to convince the wider research community that his ideas are valid, fooled him into mistaking an interesting but ultimately limited set of results for something far more significant? As his book hits the shelves, the jury is deliberating.

Precocious talent

The ideas in *A New Kind of Science* have been brewing in Wolfram's mind since the early 1980s. After ducking out of an undergraduate degree at the University of Oxford because he found it insufficiently challenging, Wolfram was in 1978 lured to the California Institute of Technology (Caltech) in Pasadena by Nobel physics prizewinner Murray Gell-Mann. The following year, he gained his PhD by submitting a bundled collection of his papers on high-energy physics and cosmology, and then joined the Caltech faculty in 1980, before moving to the Institute for Advanced Study in Princeton, New Jersey, in 1983.

"Various areas in science were stuck," Wolfram recalls. Traditional mathematical methods were, for example, struggling to

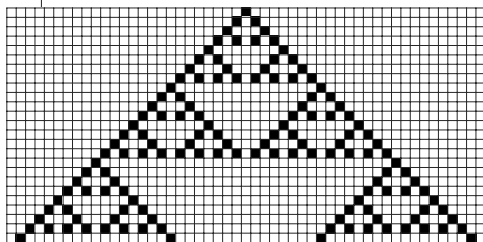
trust my judgement. I wanted to build a big intellectual structure and explain it in a coherent way.

Stephen Wolfram

show how galaxies could form from the featureless gas of the early Universe. Even mundane processes such as the growth of snowflakes were proving difficult to simulate.

Wolfram wanted to model such phenomena, and turned to simple systems known as cellular automata. The simplest 'one-dimensional' cellular automata consist of a line of squares, which can be either black or white. Each time the system is updated, a new line is created following a simple rule, and is often displayed underneath the previous line so that the evolution of the system can be tracked. One rule, for example, says that a square in the new line should only be black if one or the other, but not both, of its predecessor's neighbours were black. Starting with a single black square in the first row of squares, this rule produces a pattern of nested triangles (see left).

Cellular automata can also operate in two or three dimensions, and can use a number of different colours. But pick the right rule, and even the simplest automata can produce behaviour that is very complex or completely random. The Hungarian mathematician and



Starting with a single black square, a simple cellular automaton can generate nested triangles.

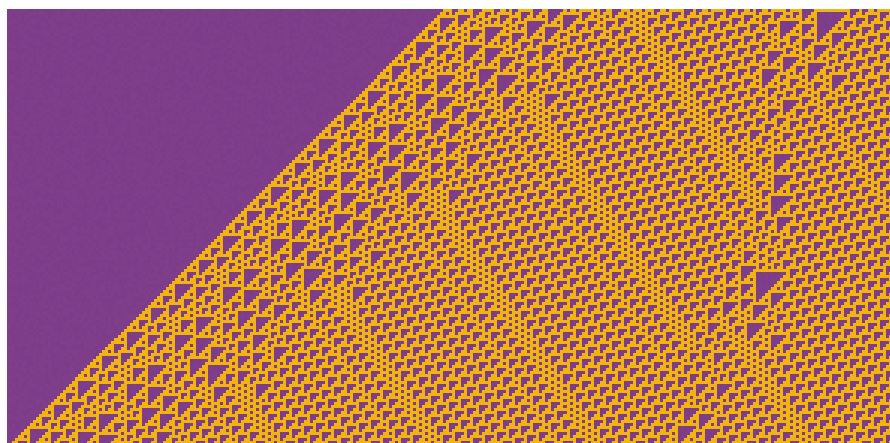
computer-science pioneer John von Neumann, a predecessor of Wolfram's at the Institute for Advanced Study, toyed with cellular automata in the 1950s. But interest had all but evaporated when Wolfram rediscovered them. "I sent my second paper on the subject to *Nature*," he says. "I got a rejection which I couldn't figure out. Then I realized it was the letter they sent to cranks."

But Wolfram persisted, and in 1984 a review article of his made *Nature*'s cover (S. Wolfram *Nature* 311, 419–424; 1984). Wolfram and a small band of other scientists and mathematicians were by then showing that cellular automata could model the behaviour of many complex systems. Snowflake growth no longer seemed so mysterious, and fluid turbulence became tractable without recourse to complicated equations. Catalysed by rapid improvements in computing power, interest in the field of complex-systems research mushroomed.

Patterns pending

Modelling using cellular automata remains an important strand within the field of complexity. But from the start, Wolfram felt that his colleagues were missing the point. "Most people were dealing with what I thought were the most mundane aspects," he says. Rather than simply using cellular automata to mimic the behaviour of complex systems, Wolfram was convinced that they could be used to reveal unknown aspects of the systems that they were modelling.

So began Wolfram's withdrawal from the academic community. His energies became increasingly devoted to perfecting software to run his cellular automata, and in 1988 he quit as head of the Center for Complex Systems Research at the University of Illinois at Urbana-Champaign, a post created for him just two years before. His new focus was the company he had founded in 1987, Wolfram Research, which was by then ready to launch the Mathematica software package. Much more than a means to run cellular automata, this program provides a convenient platform



All in one: the simple 'rule 110' can perform the same range of calculations as any physical computer.

for just about any kind of mathematical operation. Today, it is used by millions of people including scientists, engineers and financial analysts.

While Mathematica has been rising to its present dominant position, Wolfram's labour of love has remained his masterwork on the power of cellular automata and other simple systems. But it is a love that he has largely kept to himself. "Interaction slows things down," he says. Peer review is a "distraction" — indeed, Wolfram seems to think that he has few peers. Just a select few academic friends have been consulted on an occasional basis.

"I trust my judgement," says Wolfram. "I wanted to build a big intellectual structure and explain it in a coherent way. A stand-alone book is the only way to do so." He has even used his own company to publish the volume. "It harks back to the days of gentleman scientists publishing at their own expense," observes Gregory Chaitin, a mathematician and computer scientist at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, one of the few to have been consulted by Wolfram.

Researchers who have seen the book describe it as provocative and exciting, and some believe it will influence future work in their fields. But others are asking how much is really new, and suggest that Wolfram's enthusiasm for cellular automata has got the better of him.

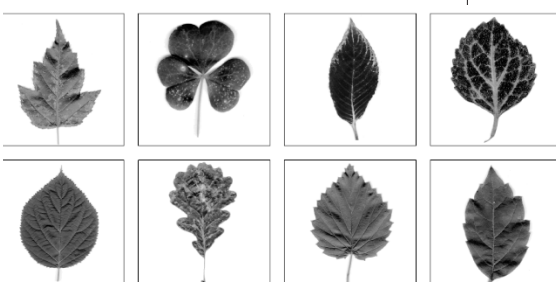
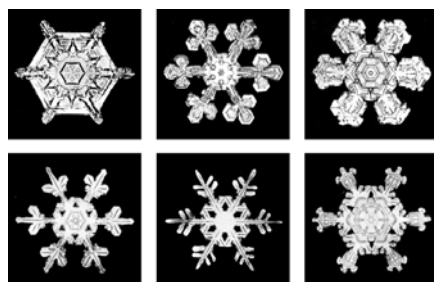
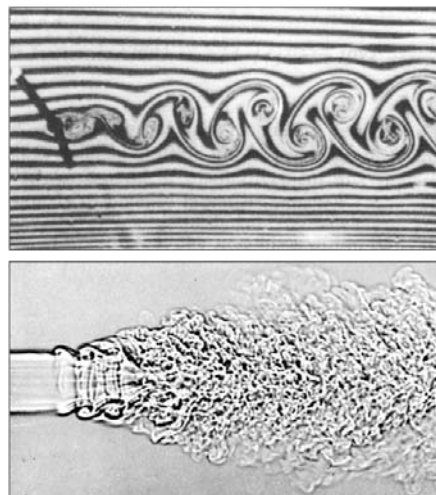
The book describes the behaviour of thousands of different cellular automata

and other simple rules, numbered according to a logical scheme. It argues that these systems can yield important insights into phenomena from biological evolution to the fundamental laws of physics. But extraordinary claims demand extraordinary evidence — which many experts feel that Wolfram fails to provide.

Everyone who has read the book says it contains some fascinating nuggets. One result, concerning a system called rule 110, is of particular interest. Rule 110 is very simple — it is one dimensional, uses only two colours, and each square can be updated by looking at just three squares in the previous row. Like all cellular automata, it can be thought of as a computer. The first line is the input, and new outputs are produced after every update. If the squares of one colour are seen as ones and those of the other as zeros, rule 110 can be thought of as doing calculations using binary numbers.

Complexity rules!

Wolfram's book shows that the results of a huge number of possible calculations lie hidden within the output of rule 110, such as computations of natural logarithms and the solutions of differential equations. In the jargon of mathematics, rule 110 is a 'universal computer' — it can perform the same range of calculations as any physical machine. More complex cellular automata have previously been shown to act as universal computers, but Chaitin and other experts are impressed with the demonstration that



Cellular automata can model a range of phenomena including turbulence, snowflakes and leaves.

very simple cellular automata can behave in the same way.

Terry Sejnowski, a computational neuroscientist at the Salk Institute for Biological Studies in La Jolla, California, another Wolfram confidant, says that the book has made him change the way he thinks about his work. Sejnowski is working on a computer model of a synapse, the gap across which two nerve cells communicate using chemical signals. He had originally entered the positions of key components of the cells by hand, but is now considering modelling the processes by which the components assemble themselves in a living cell, something he previously believed would be too difficult to simulate. "Stephen's book made me ask how the geometry of the cell arises," says Sejnowski. "He has shown this can come from simple rules. Now I'm looking for them."

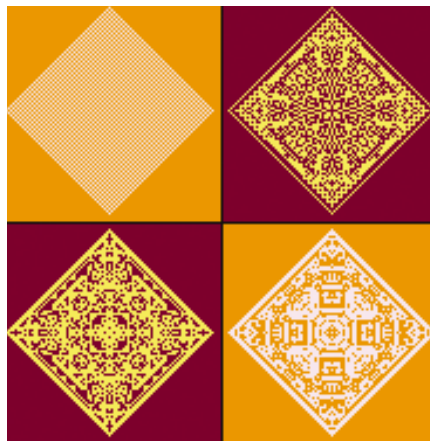
Cell divisions

Others are impressed by the book's scope, even if they disagree with some of its conclusions. Gene Stanley, a physicist at Boston University, has used other mathematical methods to study some of the same systems that Wolfram considers in his text. Stanley does not believe that cellular automata can do everything that Wolfram ascribes to them, but says that the book has persuaded him that they are more than just a curiosity. "This is a much-needed complementary approach," he says. "It's a profound book, perhaps the book of the decade."

But many experts take issue with Wolfram's expansive claims. In the section on fundamental physics, for instance, he presents a simple system, not unlike a cellular automaton, that he believes could be used to describe the fundamental basis of space and time. Wolfram argues that at extremely small scales, space is made up of discrete units, and describes a rule for determining how a structure made up of these units might evolve.

He has tested large numbers of similar models to see which produce the 'right' kind of space — one that is three-dimensional, and obeys Albert Einstein's general theory of relativity. In his book, Wolfram claims that he already has rules that do this, but admits that they cannot yet make the predictions that are possible with Einstein's equations.

Wolfram says that he has deliberately left many details to be pinned down. "I want to see the basic science take root and get a life of its own," he says. Having published the book, he is now planning an evangelical tour of research institutes and universities. "The most important thing now is education," he says. "I want to allow people to use the stuff in the book to do research. Software is coming



Curiouser and curiouser: Wolfram's book describes the evolution of a variety of automata.

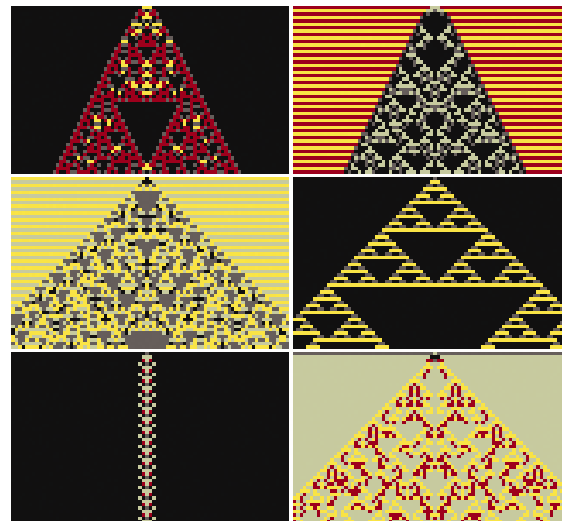
that allows people to do this." Wolfram the entrepreneur, it seems, operates hand-in-hand with Wolfram the scientific visionary.

But to many, the fact that Wolfram's ideas still lack the predictive power of established theories built on more conventional mathematics is a sign that the wunderkind has come up short. With the book's publication date having been repeatedly pushed back, some speculate that Wolfram has been striving, but never quite succeeding, to pull off his promised scientific revolution. Michael Berry, a theoretical physicist at the University of Bristol, UK, remains unconvinced that Wolfram has done more than embellish the

basic idea that simple systems such as cellular automata can generate complexity. "We've known this for 20 years," says Berry. "He'll have some fans, but I think others are going to react strongly against him."

Many in the field of complexity are already queuing up to do so. "I'm very sceptical about whether this is really a whole new way of doing things," says Doyne Farmer, a theoretical physicist at the Santa Fe Institute in New Mexico, the spiritual home of complexity research. Even the rule 110 proof has failed to set the field alight. "Lots of people are showing that all sorts of things are universal computers," says Melanie Mitchell, who works on complex systems and artificial intelligence at Santa Fe. Others point out that many of the phenomena considered by Wolfram have been modelled by other means — and are annoyed by his dismissal of rival approaches.

But within the world of complex systems it is difficult to separate reactions to the man from those to his ideas. One incident in particular has driven a wedge between Wolfram and his former colleagues. The rule 110 proof was actually developed by Matthew Cook, a young mathematician who worked for Wolfram between 1991 and 1998. After leav-



ing Wolfram's employ, Cook presented his results at a conference at the Santa Fe Institute. But details of the talk never made it into the conference proceedings. Wolfram took legal action, arguing that Cook was in breach of agreements that prevented him from publishing until Wolfram's book came out.

Difficult interactions

"We sympathized with Matthew," says one Santa Fe researcher. "Wolfram took a privatized view of science." Cook, now a graduate student at Caltech, says he cannot discuss the matter for legal reasons. Wolfram is similarly reticent — when pressed he describes the incident as "regrettable and best forgotten".

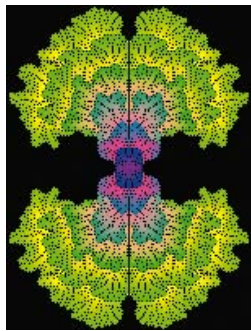
It is not the first time that Wolfram has annoyed complexity researchers, who feel that he routinely fails to recognize the contributions made by others. "He tends to acknowledge people in two-point type," says one researcher. Indeed, *A New Kind of Science* lacks conventional references to prior work — although scientists and mathematicians including Cook are acknowledged in the book's notes section.

Now that the book has finally appeared, Wolfram says that he is looking forward to engaging with his supporters and critics. "I don't want to be a recluse for another 10 years," he says. In the Boston area, from where he is promoting the book, his arrival back on the scene is causing a minor stir. Our meal was interrupted on three occasions. William Hearst, the venture capitalist who inherited the Hearst publishing fortune, popped over to say hello, as did a couple of academics, one from Harvard, the other from the Massachusetts Institute of Technology.

But Wolfram clearly desires more than his current intellectual celebrity. *A New Kind of Science* is his bid for greatness. Now all he has to do is convince a sceptical world that he is really on to something. ■

Jim Giles is *Nature's* assistant news and features editor.

♦ www.stephenwolfram.com



St Luke's new coat

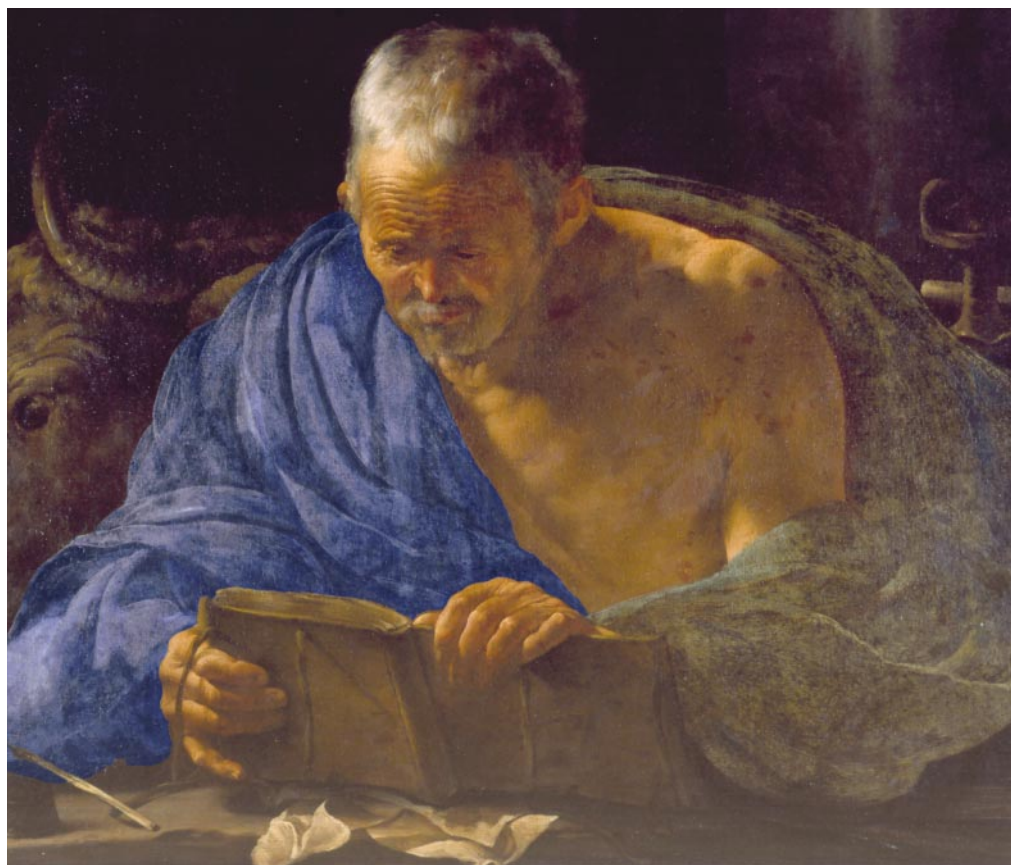
By combining chemical analyses with digital image reconstruction, it is possible to complete virtual restorations of deteriorated paintings. Celeste Biever considers the technology's potential.

When the Dutch master Hendrick ter Brugghen embarked on his *Four Evangelists*, a series of paintings that depict the authors of the New Testament, he gave each a different-coloured cloak, or mantle. Little did he know that the blue pigment he chose for St Luke's garment would not stand the test of time.

Painted in 1621, the four artworks reside in the Deventer Municipal Museum in the Netherlands. Whereas the mantles of Matthew, Mark and John are still resplendent in evocatively textured shades of yellow, red and green, respectively, Luke's cloak has long since faded to a dull and flat grey. But now Dutch researchers led by Joris Dik, a chemist at the University of Amsterdam, have produced a likeness of the original using a technique called autoradiography, combined with digital image reconstruction.

Digital reconstructions of artworks based on chemical analyses are still rare, and Dik's is the first to be achieved without any destructive sampling of the painting. Although there is still some controversy among art historians as to the validity of such techniques, some are now starting to sit up and take notice. "The possibility of digital restoration is interesting," says Martin Kemp, an art historian at the University of Oxford, UK, and a regular contributor to *Nature's Science in Culture* column.

St Luke's mantle faded because ter Brugghen used the 'smalt' pigment, a crushed, potassium-enriched glass containing silicon, oxygen, arsenic and cobalt — the latter of



Feeling blue: Hendrick ter Brugghen's *St Luke from the Four Evangelists*, seen here with one half of his mantle in its current faded state and the other in digitally reconstructed splendour.

which imparts a deep blue colour. It was favoured by artists throughout the sixteenth, seventeenth and eighteenth centuries as a cheaper, synthetic alternative to ultramarine. The latter was made from a brilliant blue mineral called lapis lazuli that had to be imported from Afghanistan.

Although art historians have known that smalt fades over time, the chemistry involved has only recently become clear. "It is difficult to get at the reaction mechanism," explains Dik. "Although it may have dramatic optical consequences, the chemical change is small."

But last year, researchers led by Jaap Boon, an analytical chemist at the Institute for Atomic and Molecular Physics in Amsterdam, showed that the fading of smalt is due to the leaching of potassium ions from the glass¹. Boon argues that the presence of potassium ions in the smalt glass encourages



Joris Dik: revealing artists' true colours.

a tetrahedral arrangement of oxygen ions around each cobalt ion. Cobalt ions can absorb a specific amount of energy to boost an electron to a higher energy level. When the oxygen ions are in a tetrahedral coordination, this corresponds to the energy of red and green light,

so this is absorbed, leaving blue light to be reflected.

But when the soluble potassium ions leach out, smaller protons replace them. Without the potassium to enforce the tetrahedral coordination, the oxygen ions surrounding each cobalt ion arrange themselves octahedrally instead, altering the amount of energy that the cobalt can absorb. This energy no longer corresponds to visible light, and the blue colour is not selectively reflected.

Colour coordination

Although they lose their ability to create colour, the cobalt ions, and traces of arsenic, remain in the ground glass. And it was by detecting their presence that Dik and his colleagues were able to attempt their reconstruction, giving a quantitative indication of where and how intense ter Brugghen intended the blue colour to be.

Dik's team put the painting in a nuclear reactor for 10 minutes to bombard it with neutrons. A tiny proportion of the cobalt and arsenic nuclei were able to capture neutrons and become unstable, radioactive isotopes — cobalt-60 and arsenic-76. These then decayed to stable, non-radioactive nickel-60 and selenium-76, respectively, by emitting β - and γ -radiation.

Radioactive arsenic-76 has a half-life of just over 24 hours. The day after irradiation it was therefore the major emitter and so



In the pink: Campaña's *The Conversion of the Magdalen* before (top) and after reconstruction.

the researchers placed a photographic film against the radioactive painting for two and a half hours to obtain a density map. Cobalt-60 has a much longer half-life, so the team applied a separate film after 10 days and left it in place for 4 months. Combining the films gave a two-tone map of where, and how intensely, cobalt and arsenic were originally applied to the painting in the smalt. The photographic films show "precisely every brush stroke", says Dik.

Hidden talents

Autoradiography has been used on paintings since the 1970s, mainly to gain insight into artists' working methods. Scientists at the Metropolitan Museum of Art in New York, for instance, have used the method to reveal Rembrandt's use of different palettes for separate sections of his paintings², and have discovered a hidden self-portrait beneath Anthony Van Dyck's *Saint Rosalie Interceding for the Plague-stricken of Palermo*². Dik's new twist is to use his autoradiographs to complete a digital reconstruction.

Dik mixed his own smalt paint according to seventeenth-century recipes. Reflectance spectroscopy of this paint allowed computer software to reproduce the colour and combine it with the two-tone radiographic mappings. Dik superimposed the resulting blue mask on a digital scan of the original painting to give the final reconstruction³.

Other experts in the chemistry of art are

intrigued by the results, although there is a lively debate about the accuracy of Dik's reconstruction. "This method is very interesting because it is the first time a digital technique has been used quantitatively," says David Saunders of the scientific department at the National Gallery in London. "However, there is one snag. It is not necessarily correct to assume the intensity of colour to be proportional to the concentrations of cobalt and arsenic."

Dik agrees, but explains that this simplification is unavoidable. The radiation signal comes from the entire paint layer but the colour you see when you look at a painting comes mainly from the upper sections. This could explain why Ernst van de Wetering, an art historian at the University of Amsterdam, finds Dik's approximation of the mantle colour too "garish".

Other groups have also recently attempted digital reconstructions. Saunders' department, for instance, has produced a reconstruction of Pedro Campaña's *The Conversion of the Magdalen*, which dates from the latter half of the sixteenth century. As a result of the instability of the red lake, smalt and green verdigris pigments, the balance of colour and some of the intended meaning has been lost. For example, Christ no longer seems to be the focus of the painting because his robe has faded from a strong pink to a pale one.

Saunders and his colleagues took pinprick-sized samples from the painting and

studied the individual pigment grains under a microscope. By supplementing this with non-intrusive, microscopic analysis, peering through cracks in the paint to deduce the layer structure, they were able to identify the pigments used and assess their relative proportions. The team then mixed their own paints according to the results of the analyses, determined their colours, and then used the results to restore the painting digitally⁴.

On the spot

But the problem with taking measurements on spot samples is that the resulting reconstruction is based on the assumption that the samples' composition is representative of a much larger area. "You have no idea how representative a sample is," says Thomas Learner, a conservation scientist at Tate Britain in London. Saunders agrees, and adds that Dik's approach has further appeal because some paintings are too precious to subject to destructive sampling. "There are some paintings that we would deeply love to sample that we wouldn't dream of attacking," he says.

Learner laments that his team cannot emulate Dik's work because he lacks the funds to gain access to a research reactor. Indeed, Dutch scientists are leading the way at the interface between chemistry and art, not least because of an initiative called MOLART, backed by the NWO, the Netherlands Organization for Scientific Research. Since 1995, more than 70 scientists interested in art, including Boon, have worked to understand the molecular basis of ageing in paintings. Their work has been funded to the tune of 7 million guilders (US\$2.9 million), mostly by the NWO. "They are my biggest competitor," jokes Dik, whose work was supported by the Nuclear Research Group in Petten and by the Amsterdam branch of Christie's auction house.

Despite the efforts of MOLART and Dik's group, some experts believe that many art historians still aren't fully aware of what science can offer. "There is still a tendency among art historians to assume that what they see now is what the artist painted," says Joyce Townsend, a senior conservation scientist at Tate Britain.

Dik and his colleagues hope that digital reconstructions will help to raise awareness of the contribution that chemistry can make, and have plans to analyse about 10 more paintings. "Autoradiography is a great tool," says Dik. "It means you can do digital reconstruction in a much more scientifically verifiable way."

Celeste Biever is currently studying journalism at City University in London.

1. Boon, J. J., Keune, K., van der Weerd, J., Geldof, M. & van Asperen de Boer, J. R. *J. Chimia* **55**, 952–960 (2001).
2. Ainsworth, M. W. (ed.) *Art and Autoradiography: Insights into the Genesis of Paintings by Rembrandt, van Dyck, and Vermeer* (Metropolitan Mus. Art, New York, 1982).
3. Dik, J. et al. *Z. Kunsttechnol. Konserv.* (in the press).
4. Richardson, C., Saunders, D. & Spring, M. *Z. Kunsttechnol. Konserv.* **15**, 338–346 (2001).

Better communication is in everyone's interests

Scientists will find it's worth making the effort to explain what they're doing to the public.

Sir — If science communicators, as characterized by Steve Fuller in Correspondence ("Communication should not be left to scientists"; *Nature* **416**, 475; 2002), are members of a "mediating profession", then teachers must surely be too.

Science reporters and producers — writing or broadcasting stories about science — provide most of society with its adult science education. They are intermediaries between scientists and the public, two groups who rarely meet face to face. We institutional press and public-affairs officers, as you rightly suggest in your Opinion article in the same issue (*Nature* **416**, 461; 2002), often provide the conduit for communication between scientists and science writers and editors.

Effective science communication requires a jack-of-all-tradesmanship uncommon in other, more specialized fields. Science communicators need to understand and interpret the scientific process and its fruits; to grasp at least the rudiments of myriad disparate disciplines of study; to translate scientific jargon into language a lay audience can understand; and to persuade editors and producers that stories warrant space or time. We are like diplomats, deftly shuttling back and forth between foreign cultures.

I was not formally trained in the sciences. If scientists can explain their work to me such that I understand it, and even go a step further and imbue me with their enthusiasm for the research, I can develop a story that will interest, intrigue and inform editors and, in turn, their readers, viewers or listeners.

This is a collaborative process in which the scientist not only submits to an interview, but also reviews drafts for accuracy. It is a time-consuming process: for the scientist, this means time taken away from his or her research. Scientists might ask whether it is worth it, and what's in it for them.

The ready answer is: recognition in the form of favourable publicity, with increased chances of funding. A more high-minded response could be the importance to civilized societies of a well-informed populace.

Public-relations professionals — science communicators among them — promote their clients' work rather than their own, which may be why we're not widely appreciated. Derisive nicknames like 'spin doctor' derive from the politicians you aptly credit with media savvy in your Opinion article. 'Spin' refers to nothing more nefarious than

the angle from which a story is told to convey the desired message.

Among other misperceptions about public relations is the belief that the practice requires no special skills and that anybody can do it. I once heard the president of a prominent US public-relations company assure his employees that what they do "is not brain surgery" — nevertheless, it is not child's play either.

The prerequisite to successful professional communication is the ability to communicate well and creatively, in speaking and writing. Understanding how the media work — including knowing the sort of story most likely to be published, being sensitive to deadlines and using tactics intended to make journalists' jobs easier — is key, as is the ability to apply strategies for managing crises and controversies.

By following your recommendation that they take a more active role in telling the "real" story, scientists will help to solve some of the genuine problems besetting science communication.

For instance, most of what constitutes science coverage in major newspapers and consumer magazines are medical and health stories. (All the examples in your editorial are medical.) It is human nature for people to be most interested in what affects them personally. The greatest

challenge by far to science public-affairs officers is finding a way to interest editors in research that does not have an immediately apparent consequence for their readers.

"Significant breakthroughs" are the only other aspects of science widely considered to be newsworthy; hence the exaggerated claims made in some press releases. People raised in the twentieth century have learned to think of science in terms of amazing discoveries, their imaginations captivated by space exploration or developments in technology that improve their daily lives. Such an audience is less easily moved by the fine points of the disciplines within my purview, such as evolution, ecology, systematics and taxonomy. But it can be done and it can be done well, without diluting the quality or integrity of the science, if scientists construct a broader, more popular frame of reference for their research.

Science communicators are trained and experienced in this line of thinking and can help. Working together, we can make science more accessible and more highly valued by the people it informs and serves.

Elizabeth J. Tait

*Office of Public Affairs, Smithsonian Institution,
1000 Jefferson Drive, Southwest, Washington,
DC 20560-0033, USA*

Boycott of Israel? It worked for South Africa

Sir — We would like to make it clear (see Opinion, *Nature* **417**, 1; 2002) that in our petition we are calling not for a boycott of individual scientists in Israel but for a suspension of institutional links until Israel complies with UN resolutions and begins to negotiate seriously with its Arab neighbours along the lines of the Saudi and other similar peace plans.

You argue that such sanctions are ineffective and that Mr Sharon will lose no sleep over them. Nevertheless, the peaceful boycott by the world's academic and cultural communities was instrumental in ending apartheid in South Africa.

If to make our demand is to be partisan, so be it — peace and justice demand partisanship. Israeli researchers may well find this prospect unpalatable, but some in Israel have selflessly given their support. Many of the signatories of our letter are receiving torrents of hate e-mail, especially if they are Jewish, being

accused of everything from anti-semitism to support for terrorism. Such allegations conflate uncritical support for the Israeli state with Jewish identity, and are both repugnant and unjust.

Finally, we would like to point out that all authors of our petition signed in their personal capacity; institutional affiliations were given for identification purposes only.

Steven Rose*, Hilary Rose†

**Department of Biological Sciences, The Open University, Milton Keynes MK7 6AA, UK*

*†City University, 4 Lloyd Square,
London WC1X 9BA, UK*

Palestinian scientists are already restricted

Sir — Your Opinion article "Don't boycott Israel's scientists" (*Nature* **417**, 1; 2002) expresses concern that a European boycott of Israeli scientists will harm three-way collaboration involving Palestinian scientists as well. This argument would

have gained in cogency by including a Palestinian perspective.

Were there reason to believe that Palestinian scientists could participate freely in such collaborations, unimpeded by arbitrary and unpredictable curfews imposed by an occupying army, denying them access to their own laboratories, not to mention European hosts, those of us who signed the call for a moratorium — not an unlimited boycott — might not have felt compelled to take such a drastic step.

Ahmed Abbas*, Mikhael Balabane*, Emmanuel Farjoun†, Michael Harris‡, Raphael Rouquier‡, Pierre Schapira§

*Université Paris 13, 99 Avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

†Hebrew University of Jerusalem, Jerusalem 91904, Israel

‡Université Paris 7, 2 place Jussieu, 75251 Paris Cedex 05, France

§Université Paris 6, 75252 Paris Cedex 05, France

Efforts to build bridges in the Middle East

Sir — I was delighted to see the Opinion article “Don’t boycott Israel’s scientists” (*Nature* **417**, 1; 2002).

I do not pretend to be impartial: my father, his brother and sister survived the Shoah, but the entire rest of his family (more than 30 aunts, uncles and cousins) perished; I have spent each of my sabbaticals partly in Israel, have sponsored students and post-docs from Israel, and have binational science foundation grants.

I also serve on the Science Advisory Board of the Arava Institute of Environmental Studies, which attempts to simultaneously do good science and build bridges between Israel and its Arab neighbours.

I count Patrick Bateson as one of my friends and grieved to see him leading the UK scientists in the direction of a boycott.

Your Opinion article beautifully articulates the reasons for continuing to support Israeli science.

Marc Mangel

Department of Environmental Studies, 39 Natural Sciences 2, University of California, Santa Cruz, California 95064, USA

Taxonomists need better access to published data

Sir — The biodiversity community must learn from its counterparts in the physical and biomedical sciences and move towards the provision of unhindered access to its baseline data: taxonomic descriptions,

imagery, geographical and temporal distribution, and characters — molecular, morphological and behavioural (see H. C. J. Godfray’s Commentary “Challenges for taxonomy”, *Nature* **417**, 17–20; 2002).

International codes of nomenclature require taxonomic actions to be published, and the data thus made available. Yet much of the underlying information is accessible only by examination of the specimens involved, so access is in effect limited to all but a few potential users. The assertion of copyright by publishers further limits the distribution of published information.

Very few libraries around the world have the financial capacity to carry the full range of literature in which systematic results are published. To take the ants as one example, the 11,000 species were first formally described in approximately 3,800 publications — roughly 100,000 printed pages — in more than 800 serials and monographs.

As F.-T. Krell noted in Correspondence (*Nature* **415**, 957; 2002), the relevance of taxonomic publications remains high for many years. Although funding has been secured to make 80% of the ant pages accessible online within the next two years at www.antbase.org (see *Nature* **416**, 115; 2002), many recent papers are not in the public domain because of publishers’ copyright restrictions.

To chart even the 1–1.5 million “known” species of the world (E. O. Wilson, *Science* **289**, 2279; 2000) is a daunting task. International initiatives such as the Global Biodiversity Information Facility are needed to help people work effectively towards that goal, as are the development of tools from information technologies and a new cultural approach to ownership and sharing of data.

In the genomics community, authors place all sequence data in a publicly accessible depository. As a result, the data themselves can be peer-reviewed, and new areas of investigation have developed through comparison and collation of data sets.

The biodiversity and conservation communities would greatly benefit from similar provision of open access to character and distributional data. Because of space and cost constraints, many of these data are unpublished. We sorely need a mechanism to provide access to these data, along the lines of GenBank, as well as the cultural imperative to deposit data (see Godfray’s Commentary for a proposal to make taxonomy a web-based unitary discipline).

For now, it would be a tremendous benefit if publishers would make published taxonomic papers open-access, so that an equivalent to PubMed can make this

important scientific information available to the broadest possible community.

Donat Agosti*, Norman F. Johnson†

*American Museum of Natural History, New York, New York 10024-5192, USA

†Insect Collection, Ohio State University, Columbus, Ohio 43210-1220, USA

Advocacy and analysis

Sir — Roger A. Pielke, in his Commentary “Policy, politics and perspective” (*Nature* **416**, 367–368; 2002) argues that scientists should not be advocates; that research should be communicated to society through policy analysts. Unlike the advocate, who subordinates science to one narrow vision of its social implications, he says that the analyst “increases the range of alternatives available to decision-makers by clearly associating scientific results with a range of choices and outcomes”.

Pielke objects to natural scientists who simply advocate more research to help solve social problems — but he then argues something similar for his own form of science: social-science policy analysis. Pielke is right that more social-scientific research and analysis of science and its results would be beneficial. But with this argument, is he not as much an advocate as those he criticizes?

Carl Mitcham

Liberal Arts and International Studies, Colorado School of Mines, Golden, Colorado 80401, USA

Specimen collecting is still vital for research

Sir — The monumental data set assembled and analysed by Ernst Mayr and Jared Diamond in their monograph *The Birds of Northern Melanesia*, reviewed by Stephen Pruett-Jones (*Nature* **415**, 959–960; 2002), was based in part on museum skin specimens obtained during general collecting expeditions over many years.

An unwillingness to acknowledge the role of specimens in scientific research is contributing to the deterioration of avian biodiversity collections and is hindering the thorough collecting that is so necessary for future studies. It has been argued that further collecting is unnecessary even in poorly inventoried areas, despite clear evidence to the contrary.

Seminal works such as Mayr’s and Diamond’s rely on specimen foundations in museums. It is essential to acknowledge this, so that the avian specimen base can be expanded, understood and supported.

Angelo Capparella

Department of Biological Sciences, Illinois State University, Normal, Illinois 61790-4120, USA

A popular evolution

An authoritative introduction to the history of life from a respected theorist.

What Evolution Is

by Ernst Mayr

Basic Books: 2001. 336 pp. \$26 (US);

Weidenfeld & Nicolson: 2002. £14.99

Mark Ridley

Ernst Mayr is a living legend. He was born in Germany in 1904, and began research in ornithology with a field trip to pre-pacified New Guinea in 1928. After moving to the United States he was soon making classic contributions to the theory of evolution. Mayr is best known for his ideas on what biological species are and on how new species originate. On both topics, his views are radically different from Darwin's, but Mayr's views prevail (if not universally) today. They are a turning point in the history of biology.

Species are populations of interbreeding organisms, argues Mayr — they are real genetic units, not artificial sets defined by phenetic attributes, or aggregate similarity. This biological species concept linked systematics with population genetics and was a key piece in the 'modern synthesis' of evolutionary biology. Mayr's theory of allopatric speciation argues that new species arise when an ancestral species is subdivided into a number of geographically separate populations. These populations then evolve independently and may over time accumulate so many genetic differences that they can no longer interbreed if they do meet.

Mayr's ideas were gathered together in his 1942 book *Systematics and the Origin of Species*, which is among the maybe five (for fans) or ten (for scoffers) most famous books ever written on evolution. He has also written standard works on ornithology, texts on systematics, a huge book on the history of biology, and has contributed ideas on the philosophy of science that are actively discussed in the literature.

Now he has written a popular book about evolution, *What Evolution Is*. He says he had three kinds of reader in mind: general readers who simply want to know more about evolution; people who accept evolution but are sceptical about the darwinian explanation for it; and creationists — not in the hope of "converting" them, but for them to find out more about what they are up against.

Accordingly, Mayr describes the history of life, including a full chapter on human origins, and explains how the history is reconstructed. He also explains the mechanisms of evolutionary change — such topics as natural selection and speciation. Mayr shows off his methuselan skills in the effortless way he sketches the history of the science, and its conceptual context on the large scale.



Hands of time: Ernst Mayr has spent a lifetime explaining the mechanisms of evolutionary change.

But what is more astonishing is how up-to-date he is. The history of life is a fast-moving research topic at present, mainly because of molecular research. Mayr is known to have slapped down a molecular geneticist or two in his time, and I wondered whether he would appreciate the new work. He does. He explains, straightforwardly and without ifs and buts, that molecular phylogenetics is where it's at, and summarizes the main findings. He seems to be thinking about the latest research like any graduate student.

He also argues as if he were in an egalitarian graduate seminar. The book bristles with provocative opinions, some of which only professional readers will understand. For instance, I detect a good-humoured argument with Mayr's Harvard colleague Stephen Jay Gould in the reiterated stress on gradualism. One paragraph bangs the point home, repeating the word 'gradual' seven times in five sentences.

Elsewhere, Mayr discusses adaptation, and asserts that "it is irrelevant for the classification of a trait as an adaptation whether it had the adaptive quality from the beginning, like the external skeleton of an arthropod, or acquired it by a change of function, like the swimming paddle of a dolphin". Experts will suspect that Mayr is dealing with Gould and Elisabeth Vrba's concept of exaptation here, but Mayr keeps quiet about that. A beginner would not know that Mayr is taking sides rather than introducing an uncontroversial point. But Mayr does support Gould on another subject: the Cambrian explosion. He suggests that the phenetic differences between organisms were greater then and that evolution may have been embryologically more anarchic at that time.

Mayr briskly and characteristically disposes of the alternatives to his preferred

biological species concept. For example: "G. G. Simpson's so-called evolutionary species concept contains undefinable criteria and is useless in praxis." Again, for experts this is a sharp judgement, to be argued with. But beginners will have more difficulty because Mayr forgets to tell us what Simpson's species concept actually is.

Mayr's intended readers may sometimes have problems when he debates without explaining that there is a debate, or debates without explaining his opponent's position. But his views all seem to me to be arguable within the range of modern biological thought, and are entirely in order in a popular introductory book. The only exception is what he has to say on neutral evolution. I'd rate Motoo Kimura's neutral theory as among the biggest, if not the biggest, contribution to evolutionary biology in the past 35 years. Mayr dismisses it in a short paragraph, as if it were an elementary conceptual mistake. He does not make his objection completely clear. However, he objects to the definition of evolution as changes in gene frequency. Neutrality is a property of genes, and neutral drift produces changes in gene frequency. For Mayr, neutral evolution is impossible because the sorts of change produced by neutral drift are not real evolution.

The book has a few dense sentences, using concepts that are either not explained or not explained until later. And it is occasionally repetitive, with topics being explained more than once in much the same words in more than one place. A duplicative transposition across 70 pages of a section on gene duplication is the most glaring example. But readers will indulge these small problems. Almost all popular books by scientists have some abstruse material. The only alternative is to read the work of a science journalist, but

then you miss out on a mind of Mayr's calibre and influence.

Mayr has now published a major ornithological monograph and a popular book on evolution in the past year. I sense he is only just getting up to speed. My next request is for an autobiography. ■

Mark Ridley is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Tuning in to radio galaxies

The Physics of Extragalactic Radio Sources

by David S. De Young

University of Chicago Press: 2002. 450 pp. \$45

K. I. Kellermann

After the Second World War, scientists, especially in England and Australia, turned their wartime radar equipment towards the skies to follow up on the pre-war discovery by Karl Jansky and Grote Reber of radio emission from the Milky Way. Although hampered at first by poor angular resolution and limited sensitivity, they were able to detect powerful sources of non-thermal radio emission from several distant galaxies. During the 1950s and 1960s, powerful interferometric and aperture-synthesis techniques were developed. When combined with theoretical advances, primarily in the Soviet Union, they showed that the observed radio emission was due to synchrotron radiation emitted from extensive clouds of ultrarelativistic electrons moving in weak magnetic fields far removed from the parent galaxy. The energy required in the form of relativistic particles and magnetic fields appeared to be enormous.

But some extragalactic radio sources were found to be very much smaller and surprisingly appeared to vary in intensity on timescales of months or less. Interferometers and lunar occultations were used to locate radio sources accurately in the sky, leading to observations of ever more distant radio galaxies and to the discovery of quasars. By the late 1960s, tape-recording interferometers had stretched radio-interferometer baselines to intercontinental dimensions. With angular resolutions better than 0.001 arcseconds available, radio astronomers were able to show that the source of energy that powers radio galaxies and quasars was as small as a few light years across or less.

Today, radio telescopes are a million times more sensitive than the pioneering post-war instruments, and more than a million extragalactic radio sources have been catalogued. Radio astronomers use wavelengths ranging from less than a millimetre to several metres, which is comparable to the range from the



Radio star: the Parkes Radio Telescope near Alectown in Australia detects extragalactic signals.

infrared through the optical and ultraviolet to the X-ray parts of the electromagnetic spectrum. Huge radio-telescope arrays show remarkably fine detail in the synchrotron clouds, revealing thin jets that carry the relativistic electron gas from active galactic nuclei (AGN) out to distant radio lobes, as much as a million light years or more away from the central engine. Global radio-interferometer observations using angular resolutions up to 100 times better than that of the Hubble Space Telescope show directly the ejection of relativistic plasma from the central engine with apparent velocities in excess of the speed of light. This is an illusion caused by the finite signal-propagation time from a relativistically moving source of radiation.

However, the new observations may have raised more questions than they answer. The source of energy is still unclear, although it is widely assumed to be associated with accretion of material into a huge black hole up to 10^9 times as massive as the Sun. How the energy is converted into a highly collimated beam of relativistic particles remains a mystery. Why are some galaxies powerful radio sources, whereas others are not? Why are powerful quasars sometimes radio-loud and sometimes radio-quiet? To what extent is the wide range of observed extragalactic radio-source properties intrinsic, and how much is due to geometrical effects? What is the relationship between quasars, AGN and galaxy formation, especially the burst of star formation often seen in distant galaxies?

In *The Physics of Extragalactic Radio Sources* De Young discusses all of these issues, starting with a qualitative review of the properties of radio galaxies and quasars. Other chapters discuss the application of synchrotron radiation, plasma and jet physics relevant to extragalactic radio sources. Classical formulae are derived only where the author has a new approach to the derivation

that leads to a better understanding of the physics involved. In other cases, the reader is referred to the appropriate literature. In the later chapters De Young applies the theory to both extended and compact radio sources, relativistic jets and AGN, including a discussion of unified radio-source models which attempt to explain the range of observed properties as the result of the orientation of the relativistic jet or an obscuring torus surrounding the central engine. He discusses the nature of the host galaxy or associated quasar and includes a review of their optical and X-ray properties.

There are few original drawings. Most of the illustrations are taken from the literature, but are sometimes out of date in this rapidly moving field and often have inadequate captions or explanation. There are a few errors, some of which seem to have been introduced by the editors. Nevertheless, the book will be a valuable resource for all astrophysics students and to everyone working in the field of extragalactic astronomy. As the author points out in the preface, no other book is devoted to this important field of astrophysics. ■

K. I. Kellermann is at the National Radio Astronomy Observatory, Edgemont Road, Virginia 22903, USA.

Exactly the same but different

The Secret

by Eva Hoffman

Secker & Warburg: 2001. 263 pages. £15.99

Justine Burley

One frequently voiced objection to human cloning is that clones would not have the same opportunities for individual development as children produced by sexual reproduction (see, for example, *Nature* **412**,

583; 2001). In *The Secret*, Eva Hoffman's first work of fiction, she probes the psychological impact on a cloned child, and on others around her, of the knowledge of her origins.

The story is narrated by Iris, the cloned daughter of a beautiful, highly successful woman (the cell donor) who has retreated to a leafy midwestern US suburb to bring up her child away from negative influences and prying eyes. Iris's cloistered childhood is characterized by a deep sense of being at one with her mother — this feels natural to both but discomfites everyone else. It is also marked by *The Weirdness*, or ill-defined self-knowledge, and *The Strange Look*, which Iris detects in others, including her mother. Her mother's lover, who was unable to discover the secret of the household, leaves when Iris is in her teens, intensifying her curiosity about her 'real father'. She snoops and discovers that she is a clone.

Iris's now rational awareness of her beginnings pitches her into an emotional void and, after violently rejecting her mother, she sets off in search of her self. Iris's sense of anomie intensifies when relatives from whom she craves love and acceptance are unable to respond to this identical copy of their sister/daughter. Desperate yet defiant, Iris sleeps with her mother's ex-lover. She does ultimately carve out a life for herself but it is devoid of family contact — her identity has been won at great cost.

The message that will be drawn from *The Secret*, whether or not it was intended

by Hoffman, is that cloning has adverse, acute and disturbing effects on the clone's personal development and on family and other social relationships, and should therefore be banned.

But would all instances of human cloning entail deleterious psychological burdens and dysfunctional relationships of the kind described by the author? If Iris had been provided with information in childhood about how she was conceived, would learning that she is a clone be so traumatic? Might not a clone instead feel special in the same way that children born by *in vitro* fertilization reportedly do? If Iris's mother had not been so overprotective, would her child have felt less isolated from, and less different to, her peers? If her extended family had not been opposed to cloning, could she not have had satisfying family relationships in which her own personality was recognizable from early infancy and in which her physiological similarity to her mother preoccupied others less? Do we not have good reason to be suspicious of any man who cannot adequately distinguish emotionally and sexually between his ex-partner and her daughter, whatever 'sameness' exists between the two women? The answers invited by these questions suggest that Iris's experience was not necessarily an inevitable consequence of her cloning origin.

Even if we assume for the sake of argument that all clones would be likely

to face the kind of struggles that Iris does in *The Secret*, is this enough of a reason to ban human cloning? The answer has to be no. If the reasonable prediction of psychological damage, the compromise of autonomy and fraught relationships is weighty enough to support a ban on asexual reproduction, then it must also be weighty enough to support a ban on sexual reproduction. It is mistaken to say that cloning involves harm of a different magnitude, because it can be seen that the welfare of children born by natural procreation is compromised severely and permanently by various behaviours of their parents, as well as by societal attitudes. It may be that no one should be allowed to procreate when it is likely that their offspring will suffer welfare-related harms of the kind discussed. But few people are prepared to stand up and argue this.

Hoffman uses the vehicle of fiction to explore a serious topic. Her strengths as an author lie in her ability to describe dislocation, alienation and individuality against different backdrops; see, for example, her much-acclaimed autobiography, *Lost in Translation*, in which, as a Polish emigrant, she struggles to find a sense of self and place in the New World. However, readers of *The Secret* should exercise caution if they were hoping for a comprehensive psychosocial commentary on the impact of cloning. ■

Justine Burley is at Exeter College, University of Oxford, Oxford OX1 3DP, UK.

Science in culture

Cosmic opera

Poussières d'Etoiles (Stardust), a multisensory experience in Paris.

Alison Abbott

Technology and imagination act symbiotically in the 'cosmic opera' *Poussières d'Etoiles* (Stardust). Conceived by Philippe Corbin and designed for the cavernous hall of the Cité des Sciences on the outskirts of Paris, it addresses the senses of sight, hearing and smell. With few words, the three-act, anti-anthropocentric show tells us that we are made up of atoms formed in stars that may have died before our Sun was born.

Spectators walk into a complex light show that takes up all 50,000 cubic metres of the hall, and are greeted by a smell of damp earth and sounds that move from pure frequencies to scratching, chattering electronic music. Light images of immense butterflies take wing. The situation is intended to place the spectators outside their normal dimensions and prepare them for their virtual journey from the Big Bang to the creation of life on Earth.

The second act begins with the activation of escalators that carry spectators to luminous sofas on the mezzanine level. Here they lie and



Eye of the beholder: the second act of *Poussières d'Etoiles* shows the development of the Universe.

watch, for 30 absorbing minutes, a breathtaking series of images projected onto the 2,400-square-metre ceiling. Many of the images are from the Hubble Space Telescope. Astronomers may recognize the Horsehead nebula, the Hubble Towers (interstellar clouds) and the Pleiades star cluster, which go by without comment as the Universe develops.

The music, like the Universe, becomes more structured, eventually taking the form of an operatic libretto, although without significant text. The composer, Nicolas Frize, says that the celestial bodies have neither soul nor feeling, so

literal dialogue in a libretto would be out of place.

For the final act, after the Earth has made its first appearance in the virtual heavens, the voices of the choir evolve into drops of water, then torrential rain. The spectators look down the escalator and see the hall submerging in virtual floods as the Earth's primeval oceans fill. They descend to watch the first organic molecules appearing in the oceans, then see life emerging in all its forms. The smell of perfumed humidity, like a sauna, pervades.

The glory of the show lies in the light-handed presentation of the beauty of astronomy and cosmology, which avoids didacticism. Frize's intelligent, evocative and moving music adds fundamentally to the poetry of the second act. But the show's weakness is its clumsy choreography of spectators, who are herded around like confused package tourists, often understanding what they should be experiencing only after the moment has passed. ■

Alison Abbott is *Nature's* Senior European Correspondent.

Poussières d'Etoiles runs nightly, from Tuesday to Saturday, at the Cité des Sciences in Paris until 2004.

Selfish drivers

Steven Henikoff and Harmit S. Malik

We are entering the post-genomic era, in which complete catalogues of genetic sequences will, it is widely believed, greatly advance our understanding of biological function. Yet there are large regions of genomes (the complete genetic catalogues of organisms) for which sequencing has hardly begun. Ironically, these are the same regions that were the first to be functionally characterized. As long ago as 1880, cytologists recognized that each chromosome in a cell nucleus has a single region, the centromere, that is acted upon by the mitotic spindle, a fibrous network that pulls sister chromosomes to opposite poles at the beginning of cell division. Thus, even before it was realized that the chromosomes' arms contain the stuff of inheritance, the role of centromeres in chromosome segregation was understood.

Molecular characterization of centromeres has been slow because of technical difficulties, not because of any lack of appreciation of the importance of these structures in cell division. Centromeric DNA in complex genomes consists of tandem repeats, which are difficult to clone, amplify or sequence. Centromeric cores are the most uniformly repetitive regions, consisting of megabase-sized, homogeneous arrays of short tandem-repeat units. These are flanked by more heterogeneous repeats, including the descendants of transposable elements — sequences that can insert themselves randomly throughout genomes. The flanking repeats mediate cohesion between sister chromosomes, and simultaneous dissolution of cohesion on all chromosomes defines the beginning of the anaphase of mitosis, when sister chromosomes begin moving to opposite spindle poles.

This functional distinction between the centromere's core, which is where the spindle microtubules attach, and its flanks, which maintain cohesion, mirrors a biochemical distinction. Half of our genetic material by weight is in the form of octamers composed of four histone proteins (H2A, H2B, H3 and H4), around which DNA is wrapped to form nucleosomes, the basic units of chromatin. Centromeric cores consist of nucleosomes with a variant H3 histone (CenH3) that replaces the usual version. Flanking sequences are packaged into 'heterochromatin', which is distinguished from gene-rich 'euchromatin' by the presence of methyl groups attached to lysine at position 9 of H3. Just how differences in H3 specify chromatin function is of substantial current interest.

Considering that centromeric cores are universally conserved in function, it seems paradoxical that their basic structural components are evolving rapidly. But centromeric repeats indeed undergo rapid evolution, evidently because of frequent homogenization events within centromeric cores. As a result, centromeric repeats differ between closely related species. Rapid evolution also occurs in CenH3s, a surprising observation given that conventional H3 is one of the most highly conserved proteins known. Indeed, this rapid evolution in both plants and animals is adaptive, apparently in response to their rapidly changing cores. Unlike other nucleoprotein machines, such as ribosomes, in which sequence conservation reflects conserved function, centromeric DNA and CenH3s continually diverge in sequence but do not vary in function. Why?

Adaptive evolution of proteins implies an 'arms race', as typified by host-parasite interactions. For example, a host immune system attacking a virus's coat protein selects for mutation of the coat protein to a resistant form, which in turn selects for mutations in the immune system that renew the attack. These adaptations result in greater rates of amino-acid replacement than would be expected in the absence of selection. In the same way, high rates of CenH3 evolution can be understood if centromeric repeats are thought of as selfish elements that constantly compete for survival, and CenH3s as host proteins that adapt in response. However, centromeres are essential for mitosis: even a single loss is highly deleterious, so that competition cannot benefit centromeres unless it involves a process by which chromosome loss does not compromise fitness.

Female meiosis is such a process. Here, the duplicated maternal and paternal centromeres separate from each other, and only one of the four meiotic products is retained

Centromeres

Centromeric DNA repeats can be thought of as selfish elements that constantly compete to beat the odds and make it into the egg at meiosis.

in the egg; the other three are degraded. There is thus the opportunity for darwinian competition between genetic variants of centromeres to reach the preferred pole and end up in the oocyte (which will form the egg). Once this one-in-four bottleneck is present, genes that reinforce it will favour their own inclusion in the oocyte if they are linked to successful centromeres.

The consequences of centromere competition would be profound. Even a slight advantage at each female meiosis will lead to rapid fixation of the winning centromere. Expanding centromeric repeats will compete during the race to reach the oocyte (known as meiotic drive). However, this has deleterious consequences. For example, misalignment of centromeres during male meiosis is thought to trigger checkpoints that interfere with spermatogenesis and cause sterility.

Driving centromeres keep winning the coin flip, and eventually the host genome will try to even the odds. CenH3s are the host factors that are in the best position to resist drive between competing centromeres. A new CenH3 allele with altered DNA-binding preference that restores meiotic parity would alleviate the deleterious effect, and thereby be driven to fixation itself. The result is an irreversible process of centromere divergence, in which both the DNA and protein components differ from their ancestors.

In this view, the sterility of interspecies hybrids that so perplexed Darwin results from a suboptimal combination of CenH3 (or another binding protein) and DNA, which would cause termination of meiosis. Darwinian competition between opposing centromeres provides a general molecular mechanism for centromere evolution that inevitably results in sterility defects in hybrids, thus accounting for the origin of species. ■

Steven Henikoff and Harmit S. Malik are at the Fred Hutchinson Cancer Research Center and the Howard Hughes Medical Institute, 1100 Fairview Avenue North, Seattle, Washington 98109, USA.

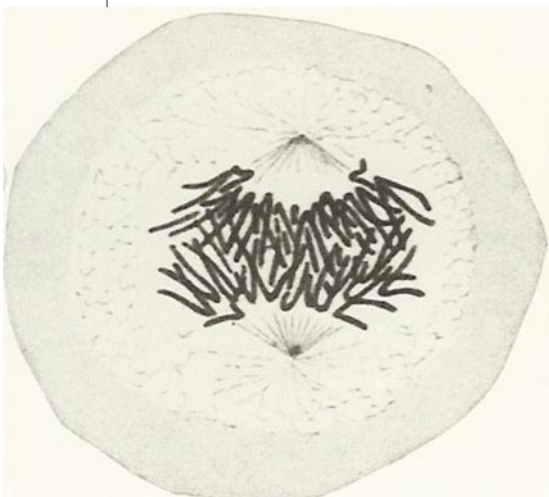
FURTHER READING

Henikoff, S., Ahmad, K. & Malik, H. S. *Science* **293**, 1098–1102 (2001).

Pardo-Manuel de Villena, F. & Sapienza, C. *Genetics* **159**, 1179–1189 (2001).

Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K. & Willard, H. F. *Science* **294**, 109–115 (2001).

Zwick, M. E., Salstrom, J. L. & Langley, C. H. *Genetics* **152**, 1605–1614 (1999).



This 1880 image shows the bulky chromosomes anchored to the spindle by their centromeres.

Thinking big in ecology

Sean Nee

A conjunction of ecology and evolution has recently produced offspring — the discipline known as macroecology. This bonny baby takes a large-scale view of the world and finds it full of contradictions.

Occasionally, the purpose of an international conference is similar to that of a baby shower — to welcome a new arrival. In this case* the infant is the discipline of macroecology, in which the world is viewed on such large scales of time and space that ecology meets evolution.

On the global scale (Fig. 1), the most significant ecological pattern is that species diversity increases towards and peaks in the tropics. Deceived by Mercator projection maps, most people are unaware that the tropics are by far the largest ecological zone, compared, for example, to tundra or boreal forest. This, surely, must be a contributing factor to the diversity gradient (M. Rosenzweig, Univ. Arizona). There are several possible explanations for why larger areas may contain more species. One version of a steady-state theory postulates that species with larger ranges may be less likely to become extinct and more likely to speciate, leading to higher diversity at the equilibrium between these processes. But some evidence does not support this connection between range size and speciation, and even suggests the opposite (A. Vogler, Natural History Museum, London; D. Jablonski, Univ. Chicago).

Looked at another way, there are good reasons to think that the latitudinal diversity gradient represents a non-equilibrium state of affairs. For example, the Arctic marine fauna was wiped out by ice ages 3–5 million years ago, and is still extremely poor in species compared to the Antarctic Ocean (A. Clarke, British Antarctic Survey). Also, many of the observed marine diversity gradients are found in the youngest taxonomic groups, implying that they are speciating in the tropical 'cradle' and spreading outwards.

Primarily, the tropics are rich in plants and their associated insects (there aren't many mammals there, apart from bats). This comparative wealth in plant species points to incoming solar energy as a likely driver of diversity. Many studies show a hump-shaped relationship between plant productivity and energy, however, with production declining at high energy levels. This problem might be resolved by the observation that water availability — an obvious, if neglected, plant requirement — also

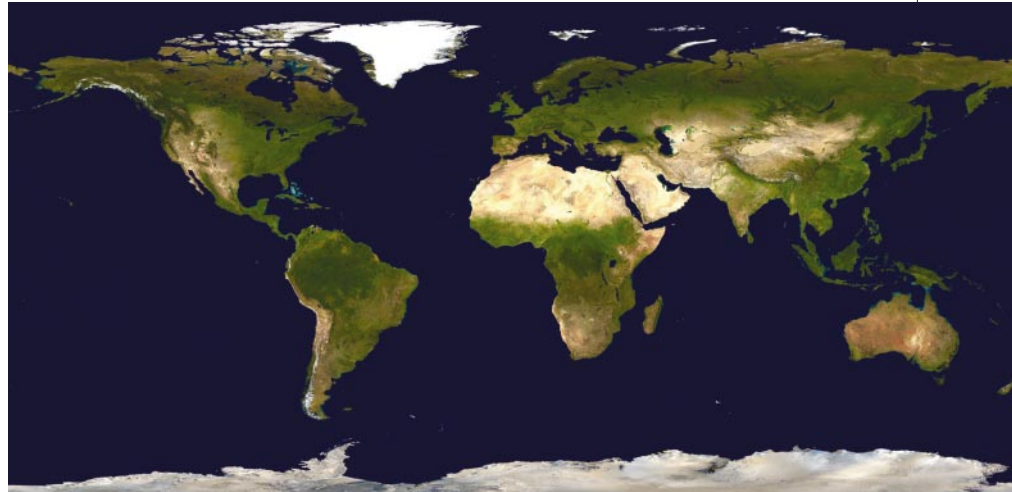


Figure 1 World view for macroecologists — and plenty of other scientists besides. This is a detail from one of the Visible Earth/Blue Marble true-colour images of the Earth, depicting every square kilometre of the planet. The full range of images draws on data from satellite observations of land, oceans, sea-ice and cloud, drawn largely from NASA's MODIS remote-sensing device on board the Terra spacecraft. This one shows a compilation of information on the land surface, ocean colour and sea ice. See ref. 4 for more details.

declines at high solar energies (R. Whittaker, Univ. Oxford).

For a journalist, a person's age is the single most informative number about them. For an ecologist, it is a species' body size, which determines the flow of energy and nutrients through individuals of that species. Indeed, understanding how metabolic rate scales with body size and, in turn, how almost all other variables scale with metabolic rate, may be the Rosetta Stone of macroecology (J. Brown, Univ. New Mexico; B. Enquist, Univ. Arizona; B. Maurer, Michigan State Univ.).

The distribution of body sizes is a fundamental topic of macroecological description and explanation. It is a basic macroecological fact that there are more small-bodied species than large ones. Strangely, this seems not to result from any propensity of smaller-bodied species to speciate: in a wide-ranging study of mammals, at no level in the evolutionary hierarchy is there a link between body size and speciation (A. Purvis, Imperial College, London). This apparent contradiction can be reconciled if two conditions are met. First, if Cretaceous mammals — that is, the earliest ones — were small (they were). Second, if, after events at the Cretaceous–Tertiary boundary, when the impact of an extraterrestrial body is thought to have caused

widespread extinctions, those mammals that survived then speciated and evolved randomly with respect to size.

Along with a large-scale view of the world, macroecology has large unifying visions. Neutral theories — in which the numbers of individuals in the species simply fluctuate randomly in time and space — have recently become prominent. These theories have successfully generated many observed macroecological patterns, correctly predicting the relative population abundances of the species in a community, for example. They also predict the observation that species with higher population densities tend to have larger ranges — thus correlating these two 'dimensions' of abundance (J. Lake, Univ. Georgia; G. Bell¹).

The fact that neutral models can generate observed patterns has two implications. First, how can we confirm or refute the neutral models? It is well known that completely different mechanisms can generate the same pattern: the distribution of parasitic worms among people is the same as the distribution of word usage in Shakespeare — the negative binomial². This means that the patterns themselves cannot inform us about mechanism and some other techniques are needed. Suitable methods need to be developed urgently.

*Macroecology: Reconciling Divergent Perspectives on Large-scale Ecological Patterns, British Ecological Society Symposium, 16–19 April, Birmingham, UK.

The second implication is closely related to the first. If a simple random model can describe a set of data, this perhaps shows that the data are at a scale of resolution at which the mechanism generating the data cannot be discerned. So, for example, the number of outbreaks of war in a given year is well described by a Poisson distribution³, but this does not mean that foreign policies are guided by the clicks of a Geiger counter monitoring poissonian radioactive decay. It should be emphasized that even if macroecological data are at such a scale, this in no way compromises their great value, especially in areas such as conservation biology. At the meeting we enjoyed the example of a macroecologist from Norwich informing New Zealand fisheries biologists that one of the species in their sphere of study was probably in trouble, and in need of monitoring (J. Reynolds, Univ. East Anglia).

In pursuit of the high purpose of consolidating a new field, serious contention was left to the future, as with the issue of slavery during the convention that drew up the constitution of the United States. Not that

contentious issues were hidden: each session was structured in accord with Hegelian dialectical philosophy, with each talk being either a thesis or antithesis. We have the following two debates to look forward to, for example. One is whether diversity has been continually increasing over the past several hundred million years, with the odd blip of a mass extinction (A. Clarke), or whether it has not, appearing instead to be at a steady state (M. Rosenzweig). The other is whether metabolic rate increases with body size in the same fashion — the well-known 3/4 power scaling — both within and between species (B. Enquist), or whether it does not (J. Kozłowski, Jagiellonian Univ., Krakow). Onwards to synthesis.

Sean Nee is at the Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.

e-mail: sean.nee@ed.ac.uk

1. Bell, G. *Science* **293**, 2413–2417 (2001).
2. Efron, B. & Thisted, R. *Biometrika* **63**, 435–447 (1976).
3. Richardson, L. F. *J. R. Stat. Soc.* **107**, 242–250 (1944).
4. <http://visibleearth.nasa.gov/cgi-bin/viewrecord?11605>

nanoparticles and *ab initio* calculations of model $\text{HCl}(\text{H}_2\text{O})_6$ clusters, the authors identify four steps of hydrogen bonding and proton transfer. These results should have an impact in areas where ‘confined’ water or ice is involved — such as proton transfer in biological channels (‘water wires’); hydrogen bonding and proton-transfer catalysis in water layers within the active sites of enzymes; atmospheric surface reactions on water droplets and proton transfers on stratospheric ice particles.

Several groups have modelled the dissociation of HCl in and on small clusters in liquid water and at the surface of ice by classical or quantum-dynamical simulations and *ab initio* calculations. By measuring the infrared frequency shifts of the surface-adsorbed HCl molecules and comparing the results to theory, Devlin *et al.*¹ have now pinpointed three configurations of HCl adsorbate on the ice surface layer. The three steps, shown in Fig. 1, are the formation of a single $\text{Cl}-\text{H}\cdots\text{O}$ hydrogen bond to a water molecule in the ice surface; the formation of a second $\text{O}-\text{H}\cdots\text{Cl}$ hydrogen bond to a nearby dangling OH group; and the formation of an additional $\text{O}-\text{H}\cdots\text{Cl}$ hydrogen bond, which is followed by dissociation of HCl to form a ‘contact ion pair’, in which the Cl^- ion remains in contact with the newly formed H_3O^+ (hydronium) ion. In the final stage of dissociation, the H^+ ion (a proton) diffuses away from the Cl^- ion along the ice surface.

The second step is particularly interesting: coordinated by two hydrogen bonds, the HCl molecule is strongly stretched, but not yet ionized. The experimental signature is a 40% decrease in the vibrational frequency of the molecule, relative to that of gas-phase HCl. Those $\text{HCl}(\text{H}_2\text{O})_6$ cluster models that yielded similar large vibrational shifts in

Physical chemistry

Acids caught in the act

Samuel Leutwyler

Using infrared spectroscopy, the separate stages of the dissociation of hydrogen chloride molecules on an ice surface have been revealed. The formation of hydrogen bonds seems to be a decisive factor in the process.

Chemical concepts have a long history of success in rationalizing the huge variety of phenomena observed at and above the molecular level. There are, for example, concepts that help to predict molecular structure — the two-electron bond, the octet rule and lone electron pairs. At a higher level, concepts such as valence-shell electron repulsion and orbital hybridization explain the preference of atoms to form a characteristic number of bonds — summed up as their valence or coordination number.

In the supramolecular world, hydrogen bonding and proton transfer are two powerful chemical concepts. The tendency of hydride molecules such as H_2O , NH_3 or HCl to form hydrogen bonds and to transfer protons is well known. But, because there is rapid thermal diffusion and reorientation of molecules in liquids, it is difficult to determine the local hydrogen-bond patterns and coordination numbers for molecules undergoing a proton-transfer reaction. But help is at hand: on page 269 of this issue, Devlin *et al.*¹ provide a fascinating view of the separate stages of hydrogen bonding during the proton-transfer reaction $\text{HCl} \rightarrow \text{H}^+ + \text{Cl}^-$ in aqueous solution.

The acid dissociation and dissolution — which is extremely rapid in solution — is

‘frozen in’ by performing the reaction at the surface of ice nanoparticles at temperatures between 50 and 90 K. Using a powerful combination of infrared difference spectroscopy, computer simulations of the ice

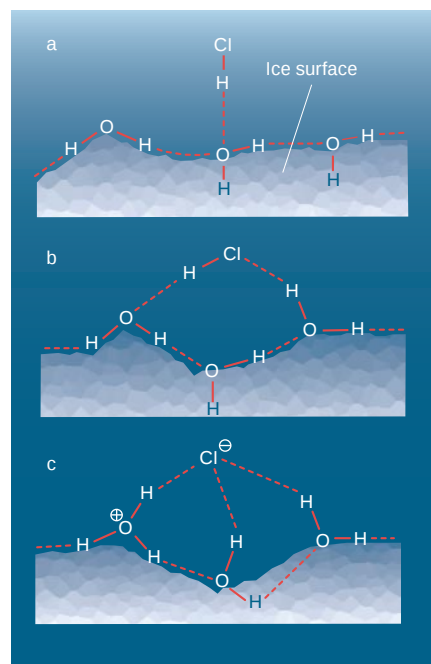


Figure 1 Dissociation of hydrogen chloride (HCl) on the surface of ice is intrinsically linked to the number of hydrogen bonds formed. Using spectroscopic analysis of a mixture of HCl and nanoparticles of ice, Devlin *et al.*¹ were able to capture the process of dissociation. a, First, a single strong hydrogen bond forms, linking the HCl molecule to an oxygen atom on the ice surface. b, A second hydrogen bond forms between the Cl atom and a ‘dangling’ OH group. c, Finally, a third hydrogen bond connects another dangling OH group to the Cl atom, followed by dissociation of HCl, leading to a ‘contact ion pair’ of chloride ion and H_3O^+ ion. The formation of three hydrogen bonds defines the ‘hydrogen-bond coordination number’ necessary for HCl dissociation. This concept could usefully be extended to understanding other proton-transfer reactions, particularly in biochemical systems.

the *ab initio* calculations confirm the double coordination for HCl. Beyond that, they reveal fascinating details of the role of the hydrogen-bond coordination of the acceptor H₂O molecule: only a recipient H₂O that has not already accepted another hydrogen bond (a two-coordinated dangling-O molecule, meaning it has formed two OH donor hydrogen bonds) is able to induce both the large vibrational frequency shift in the second step and proton transfer in the third step. An H₂O molecule that is already three-coordinated is a weaker proton acceptor than a two-coordinated dangling-O water molecule and does not induce proton transfer.

The concept of hydrogen-bond coordination numbers helps us to understand proton transport in water. The simplest view of a hydrated proton is the H₃O⁺ ion, which has an average coordination number of three. But H₂O molecules have an average coordination number of almost four. Recent theory has focused on the interplay² between the triply coordinated hydronium H₃O⁺(H₂O)₃ (the 'Eigen cation') and H₅O₂⁺ (the 'Zundel cation'). In the latter, the proton is bound between two water molecules with 'stretched' bonds. The result is a strongly shifted and wide infrared absorption band, as observed in acid hydrates and concentrated acid solutions. This characteristic continuum is evident in the spectra of Devlin *et al.*, and signals the fourth step of HCl dissociation — the surface diffusion of the proton with transient formation of the H₅O₂⁺ cation.

In the Grotthuss model of aqueous proton transport, H₃O⁺ transfers one of its protons

to a recipient H₂O molecule in its first coordination shell, which thereafter passes a different proton to a further recipient water molecule, and so on. For the recipient H₂O to become a three-coordinated H₃O⁺ ion, it has to change from the average hydrogen-bond coordination number of four to only three. Some have speculated³ that the rate-limiting step involves the breaking of an H bond to the O atom of the recipient H₂O molecule before the proton transfer. This is supported by classical and quantum molecular dynamics simulations^{2,4–6}. The need for the recipient H₂O molecule to have a lower than usual hydrogen-bond coordination number is exactly analogous to the role of the two-coordinated dangling-O molecule on the ice surface in the work of Devlin *et al.*

These new insights into the role of local hydrogen-bond coordination numbers in HCl should improve our understanding more generally of proton-transfer reactivities and the complex dynamics of biochemical systems. ■

Samuel Leutwyler is in the Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland.
e-mail: leutwyler@iac.unibe.ch

1. Devlin, J. P., Uras, N., Sadlej, J. & Buch, V. *Nature* **417**, 269–271 (2002).
2. Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. *Nature* **397**, 601–604 (1999).
3. Agmon, N. *Chem. Phys. Lett.* **244**, 456–462 (1995).
4. Ando, K. & Hynes, J. T. *J. Phys. Chem. B* **101**, 10464–10478 (1997).
5. Vuilleumier, R. & Borgis, D. J. *Chem. Phys.* **111**, 4251–4266 (1999).
6. Day, T., Schmitt, U. & Voth, G. J. *Am. Chem. Soc.* **122**, 12027–12028 (2000).

Amyloid diseases

Small drugs lead the attack

Leslie Iversen

Many human disorders — a well-known example being Alzheimer's disease — are characterized by the misfolding and aggregation of key proteins. New small-molecule drugs may help to destabilize these aggregates.

On page 254 of this issue, Pepys and colleagues¹ describe a new pharmacological approach to treating human amyloid diseases. These often devastating disorders range from type II diabetes to Alzheimer's disease, and involve the abnormal folding of usually soluble proteins into an insoluble, tightly packed shape known as a β -sheet. This causes the proteins to be deposited outside cells and form aggregates called amyloid deposits that lead to tissue damage and, ultimately, to death of the patient. Several human proteins can undergo such a transformation; in Alzheimer's disease, for instance, the accumulation in the brain of insoluble aggregates of the amyloid- β (A β) peptide — also known as 'senile plaques' — is thought to be a key

pathological event underlying the loss of brain cells².

There is interest in a variety of approaches aimed at preventing or even reversing amyloid deposition, particularly in Alzheimer's disease. One tack involves vaccination, which has seen some success in mice that were genetically modified to overexpress the human amyloid precursor protein, from which the A β peptide is produced. When these mouse 'models' of Alzheimer's disease were immunized with the A β peptide itself, the formation of senile plaques in the brain and the associated behavioural deficits were reduced³. It was proposed that a small proportion of the antibody that the animals produced in response to the vaccine passed the barrier between blood and brain and 'decorated' the A β plaques, enabling them



100 YEARS AGO

News of the terrible volcanic eruption in Martinique reached this country on Thursday last, and the details which have since become known have shown that an appalling disaster has occurred. St. Pierre, the chief commercial centre of the island, has been totally destroyed, and about thirty thousand people have perished. The eruption of Mont Pelée began on the night of Saturday, May 3, when large quantities of scoriae and volcanic ash were thrown into the surrounding country. On Monday, May 5, a stream of lava is reported to have rushed down the side of Mont Pelée, following the dry bed of a torrent, and reaching the sea, five miles from the mountain, in three minutes. When the stream met the sea the water receded 300 feet on the west coast, returning with greater strength in a large wave. Two days later, on May 8, a similar torrent of incandescent lava engulfed the town of St. Pierre.

From *Nature* 15 May 1902.

50 YEARS AGO

In these days of specialist scientific societies, and with the calendar peppered with technical conferences and seminars, the British Association is no longer the place for the announcement of major new discoveries in science. In recent years the Association has been restored to its original purpose: to be a place where scientists in different fields interpret their work to one another and to the public. It might be argued that this original purpose has now been fulfilled. The public spends some £16 million a year on civilian scientific research, quite apart from £20 million a year spent on universities: surely it is no longer necessary to arouse public interest in science; and in any event have we not the Third Programme, and a flourishing popular scientific literature, to maintain interest? This is a topic which might well be discussed by the British Association itself; but there is not much doubt what the outcome would be. For scientists are now dependent on public support in a way they never were before ... Therefore the interpretation of science to the public has now become a major activity in modern society, not to be entrusted entirely to the efforts of scientific journalists; and the British Association remains the chief instrument for this activity. Eric Ashby

From *Nature* 17 May 1952.

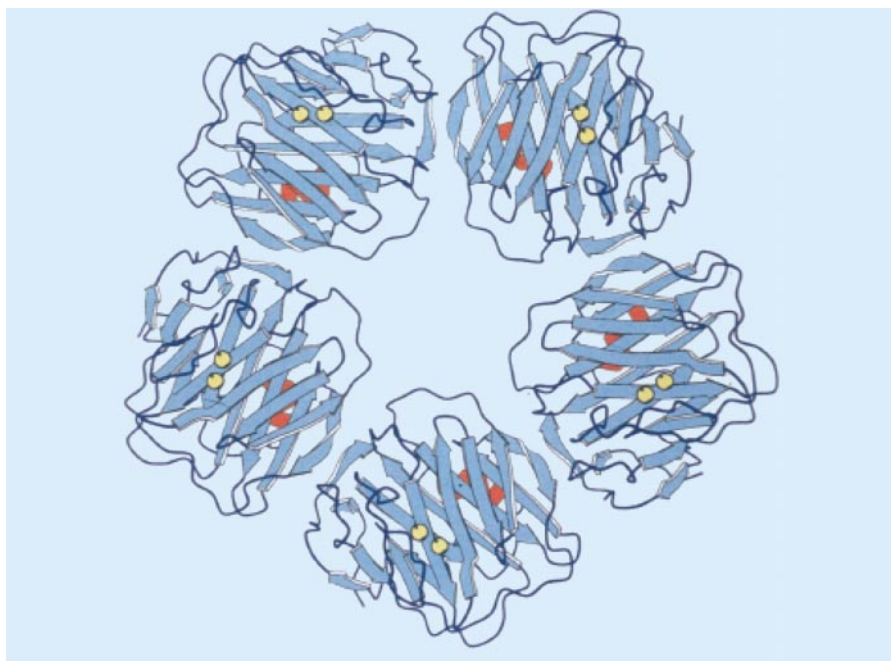


Figure 1 Structure of the pentamer of the serum amyloid P protein (SAP)¹³. Pepys *et al.*¹ have discovered some small-molecule drugs that prevent SAP from binding to amyloid deposits, and might therefore be useful in treating amyloid disorders such as Alzheimer's disease.

to be recognized and consumed by microglia (a type of white blood cell). The findings prompted clinical trials of this potential vaccine for treating Alzheimer's disease. But unfortunately, the trials had to be halted because some patients developed signs of inflammation of the central nervous system⁴. An alternative approach is 'passive' immunization, which involves injecting antibodies against the A β peptide (rather than the peptide itself); this has also proved effective in the mouse model⁵ but has yet to face clinical trials.

Another strategy involves using several low-molecular-weight substances, including dyes such as Congo red, the antibiotic rifampicin and the anthracycline 4'-iodo-4'-deoxydoxorubicin, to disrupt the aggregation of the A β peptide⁶. Screening for such compounds also identified other chemicals that are active against amyloid deposits⁶. In a related approach, short peptides that bind to the A β peptide and interfere with aggregation were synthesized⁷.

The discovery that A β deposition is accelerated by metals, notably copper and zinc, provided yet another angle of attack. The antibiotic iodochlorhydroxyquin (known by the trade name Clioquinol) chelates copper and zinc *in vitro*, and reduced A β deposition in a mouse model⁸. Moreover, interim results from a randomized, double-blind, placebo-controlled clinical trial in 32 patients with Alzheimer's disease suggested that this drug slows the rate of cognitive decline in the most severely affected group⁹.

Clinical trials are also under way with the small, sulphonated molecule NC-758 (trade name Alzhemed), which was designed to interfere with the binding of glycosamino-

glycans to the A β peptide¹⁰. (Glycosaminoglycans are polysaccharides that may help to stabilize and protect brain amyloid deposits.) NC-758 had previously proved effective at reducing A β deposition in a mouse model¹¹.

These are all imaginative strategies, but new approaches are, of course, always welcome. Pepys *et al.*¹ have come up with just that: they reveal a different way of destabilizing amyloid deposits using small-molecule drugs. Specifically, these drugs target the protein serum amyloid P (SAP), which is universally present in amyloid deposits — both in Alzheimer's disease and in peripheral amyloid disorders (which affect tissues other than the brain). SAP binds to these deposits in a calcium-dependent manner and helps to protect them from degradation by protein-degrading enzymes or white blood cells. Pepys and colleagues suggested almost 20 years ago¹² that drugs that promote the dissociation of SAP from amyloid deposits, making these deposits accessible to the body's normal removal mechanisms, might represent a new way of treating amyloid diseases. They now go a long way towards proving this concept.

The authors¹ used a high-throughput screen to test a library of chemicals for those that inhibit the binding of SAP to A β deposits *in vitro*. This led to the discovery of a series of dimeric derivatives of the amino acid proline that are highly active both *in vitro* and *in vivo*. SAP normally exists as a pentamer¹³ (Fig. 1), and the proline derivatives link two pentamers together (see Fig. 2a on page 256). This has two important effects: SAP is prevented from binding to amyloid deposits; and the drug-SAP complex is rapidly removed from the blood circulation

and degraded in the liver. Consequently, Pepys *et al.* found that levels of SAP in the blood serum of test animals were dramatically reduced. Moreover, in animal models of amyloid diseases, using the drug led to the removal of SAP from peripheral amyloid deposits, and a gradual reduction in the size of these deposits.

Pepys *et al.* also report the first results from studies in humans. One of the proline derivatives was given safely for up to nine and a half months to 19 patients with peripheral amyloid diseases. Levels of SAP in the blood plasma of these patients were markedly reduced and there was evidence that SAP levels in amyloid deposits also decreased.

So this new approach¹ offers great promise for treating both peripheral amyloid disorders and, possibly, Alzheimer's disease. The blood-brain barrier often presents a stumbling block to new drugs, but it may not be necessary for the proline derivatives to cross this barrier: lowering circulating SAP levels may be enough to deplete SAP from the brain and other tissues. Moreover, the clinical-trial data suggest that the prolonged removal of circulating SAP has no obvious harmful effects, even though it is a normal component of blood plasma. But whether lowering SAP levels will be sufficient to facilitate the disaggregation and complete removal of long-standing amyloid deposits remains to be seen.

It may be possible to extend the use of low-molecular-weight organic drugs to other protein-folding diseases. For instance, acridines and phenothiazines inhibit the conversion of the prion protein PrP^C to the disease-causing form, PrP^{Sc}, in cultured cells infected with prions¹⁴. Quinacrine and chlorpromazine have been used for many years to treat malaria and schizophrenia, and it has been proposed¹⁴ that these drugs be used in clinical trials for treating Creutzfeldt-Jakob disease. With all the activity in this research arena, we can perhaps be hopeful that drugs can be found to combat the abnormal protein deposits seen in so many distressing human diseases. ■

Leslie Iversen is at the Wolfson Centre for Age Related Diseases, King's College, London SE1 3QD, UK.

e-mail: les.iversen@kcl.ac.uk

1. Pepys, M. B. *et al.* *Nature* **417**, 254–259 (2002).
2. Selkoe, D. J. *Physiol. Rev.* **81**, 741–766 (2001).
3. Schenk, D. *et al.* *Nature* **400**, 173–177 (1999).
4. *SCRIP World Pharmaceut. News* No 2726, 23 (2002).
5. Bard, F. *et al.* *Nature Med.* **6**, 916–919 (2000).
6. Findeis, M. A. *Biochim. Biophys. Acta* **1502**, 76–84 (2000).
7. Soto, C., Saborio, G. P. & Permanne, B. *Acta Neurol. Scand.* **102** (Suppl. 176), 90–95 (2000).
8. Cherny, R. A. *et al.* *Neuron* **30**, 665–676 (2001).
9. Masters, C. in *7th Int. Geneva/Springfield Alzheimer's Symp.* 4 April 2002.
10. Garceau, D., Gurbindo, C. & Laurin, J. in *7th Int. Geneva/Springfield Alzheimer's Symp.* 4 April 2002.
11. Gervais, F. *et al.* in *7th Int. Geneva/Springfield Alzheimer's Symp.* 4 April 2002.
12. Hind, C. R., Collins, P. M., Caspi, D., Baltz, M. L. & Pepys, M. B. *Lancet* **324**, 376–378 (1984).
13. Emsley, J. *et al.* *Nature* **367**, 338–345 (1994).
14. Korth, C. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 9836–9841 (2001).

Planetary science

A wet and altered Mars

John F. Mustard

Has Mars always been cold and dry or was it once warm and wet?
Reanalysis of spacecraft data reveals a signature from the surface rocks
that indicates their composition has been altered in the presence of water.

Two years ago, Bandfield *et al.*¹ published interpretations of data from an instrument on the Mars Global Surveyor spacecraft that caused many planetary scientists to rethink their ideas about the geological evolution of the red planet. On page 263 of this issue, Wyatt and McSween² describe how they have reinterpreted the same data and arrived at a fundamentally different view. So what's going on?

Bandfield *et al.* reported that the mineralogy of volcanic regions on Mars was best explained by the existence of two very different rock types — basalt in the southern highlands and andesite in the northern plains. The mineral make-up of a volcanic rock reflects the composition of its source region and whether it has undergone melting or crystallization. On Earth, basalt (which is dominated by the minerals plagioclase and high-calcium pyroxene) is the typical rock of the ocean floor; it is formed by partial melting of mantle rocks beneath the Earth's crust and extrusion of the melt to the surface. Andesite, a more silica-rich rock, is most commonly formed by partial melting of mantle in the presence of the water released as oceanic crust descends into the mantle along subduction zones.

Bandfield and colleagues' conclusion, that there are large areas of both andesite and basalt on Mars, precipitated a number of questions. Has recycling of crust occurred on

Mars? Are there significant quantities of water in the mantle? Are the mineralogical divisions due to some other magmatic process?

Wyatt and McSween's reinterpretation² of the data leads to a very different, but no less intriguing, conclusion. Their calculations also produce a mineralogy that is consistent with basalt for the southern highlands. But for the northern plains, their inferred composition consists of typical basalt minerals plus clays and sheet silicates that are characteristic of low-temperature aqueous alteration — in other words, basalt weathered in the presence of water. It happens that the distribution of this altered basalt corresponds well with the vast areas of the northern plains in which water, debouched from enormous outflow channels that emanate from the southern highlands, may once have been concentrated into an ocean (Fig. 1)^{3,4}. The spatial coincidence is not perfect, and there are concentrations of altered basalt in the southern highlands that are not correlated with enclosed basins or water-related features. Yet the association is strong enough to take the implications seriously.

If Wyatt and McSween's solution is correct, a key question is when the basalt alteration took place. The timing would affect the specific mineralogy of the alteration, as well as the prospects for potential habitats for life. If the volcanic material was erupted into an ocean (or ice, if the ocean was

frozen), then we might expect copious hydrothermal activity to have occurred, which would have provided abundant energy sources and habitats for life. Furthermore, the alteration could have been widespread in places and penetrated deep into the crust. If, on the other hand, the oceans were created well after the volcanic rocks were in place, then without heat sources the amount of chemical exchange would have been far less. In this case, the alteration would have been relatively weak and mainly confined to the surface, and the habitats for life would have been far less favourable.

A third alternative is that the altered basaltic rocks are sedimentary in origin, with the alteration occurring at some unspecified time and location upstream from the basin, and the altered rocks subsequently being transported to their present location. In this case, the association of the altered basalt with the ocean basin would be purely a consequence of sediment transport. Certain geomorphological features in the northern plains suggest that interaction between volcanoes and water or ice may have occurred. But these features are sparse relative to the extent of alteration mapped by Wyatt and McSween². So the question of when and how the alteration occurred remains unresolved.

One might ask how such distinctly different conclusions^{1,2} can be derived from the same data set. The analyses were conducted with spectra acquired by the thermal emission spectrometer (TES) on Mars Global Surveyor, which is still in operation. Crystal-lattice vibrations in minerals modify the thermal emission spectra from surfaces in distinct ways that can be used to identify minerals and their mixtures. However, the signature from the surface of Mars is relatively weak and, compared to that from terrestrial rocks, bland. As Wyatt and McSween show², there are several possible solutions for the TES observations that fit the data equally well. Which is the right one? In fact, the solutions differ only subtly. It turns out that high-silica glass and some clay minerals have very similar spectral properties in the TES wavelength range. Bandfield *et al.* used the glass spectra, Wyatt and McSween did not. But this made a big difference to the interpretations of rock type and thus of geological history.

One source of uncertainty in these new results is that the surface of Mars is quite commonly covered to a variable extent by mobile materials such as dust and sand. True exposures of the crust are rare, although abundant rocks of likely local origin were observed at the Viking landing sites, which are in the northern plains. Poleward of 30°, both north and south, there is a patchy and discontinuous layer of cemented dust and soil, 1–10 m thick, which was formerly ice-rich⁵. This layer becomes more continuous with latitude, until it covers most of the

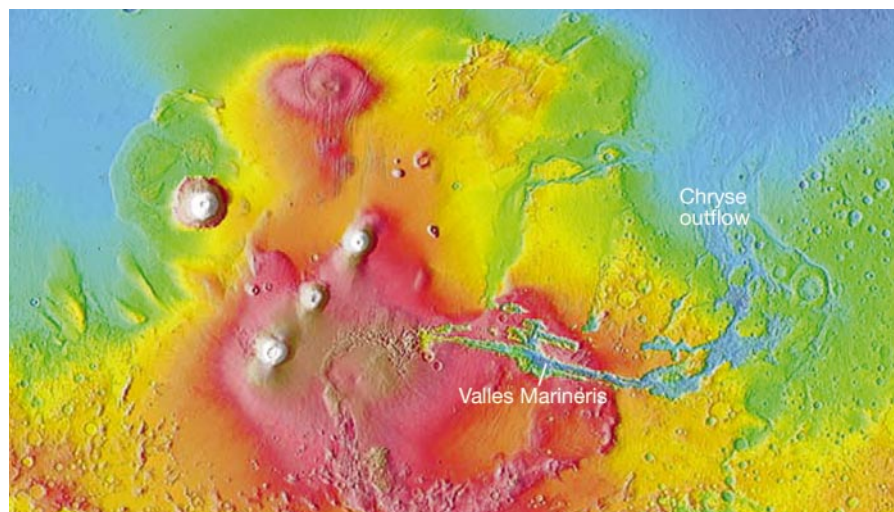


Figure 1 Mars in relief. This shaded topographic map is of the part of Mars known as the Tharsis province, and shows the major volcanoes, the Valles Marineris and the Chryse outflow regions. Colour indicates elevation, where blue is lowest and white is highest. Wyatt and McSween² report that volcanic rocks in the northern plains (blue) have been altered in the presence of water, whereas those in the southern highlands have not.

surface at latitudes above 60° (ref. 6). Water in this layer may be responsible for the alteration suggested by Wyatt and McSween, and poleward of 60° may obscure the signatures of the underlying bedrock.

To see beneath the mantle, we need to measure spectra from isolated, small exposures of bedrock, and that is not possible with the current generation of sensors, which offer only coarse spatial resolution. However, a new series of sensors, with increased spatial resolution and covering complementary wavelength regions, is scheduled to fly to Mars in 2003–2006. By combining all these

sources of data, it should be possible to identify true exposures of the martian crust and determine their mineral compositions. ■

John F. Mustard is in the Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA.

e-mail: john_mustard@brown.edu

1. Bandfield, J. L., Hamilton, V. E. & Christensen, P. R. *Science* **287**, 1626–1630 (2000).
2. Wyatt, M. B. & McSween, H. Y. Jr *Nature* **417**, 263–266 (2002).
3. Parker, T. J. et al. *J. Geophys. Res.* **E 98**, 11061–11078 (1993).
4. Head, J. W. et al. *Science* **286**, 2134–2137 (1999).
5. Mustard, J. F., Cooper, C. D. & Rifkin, M. L. *Nature* **412**, 411–414 (2001).
6. Kreslavsky, M. A. & Head, J. W. *J. Geophys. Res.* **E 105**, 26695–26711 (2000).

Cancer

New-age tumour suppressors

Allan Balmain

The discovery of a gene that is inactivated in stomach cancers illustrates the value of mouse models for finding tumour-suppressor genes that are switched off by mechanisms other than mutation.

Tumour-suppressor genes form a crucial part of our natural anticancer defences. When these genes become inactive, tumours often develop. Most cells have two copies (alleles) of every gene, and the classical view of a tumour suppressor is that loss of function only occurs when both copies are inactivated genetically. First, a mutation takes place in one allele in a developing tumour cell; next, the remaining normal allele is knocked out by 'loss of heterozygosity' (LOH) — the cell loses a large part of the chromosome on which the normal allele resides¹.

These ideas led to the discovery of a whole generation of tumour suppressors, including the well-known retinoblastoma protein and p53 (see ref. 2 for a review). However, although LOH is often seen in certain chromosomal regions in human and mouse cancers, the rate of discovery of the tumour suppressors in those regions seems to have slowed substantially. This may be because many tumour-suppressor genes take alternative routes to inactivation. Writing in *Cell*, Li *et al.*³ describe how they identified *RUNX3* — which they find to be switched off in human gastric cancers — as just such a gene.

Why have the new-generation tumour suppressors been so difficult to identify? One reason may be the strategy that is typically used: find the characteristic region of LOH in a particular tumour; narrow down that region as much as possible by 'deletion mapping'; and hunt for inactivating mutations within genes in the corresponding region on the matching, undamaged chromosome. This allows the discovery of classical tumour suppressors, which have genetic alterations in both alleles. But tumours are more resourceful than we once thought, and as well as genetic mechanisms, they use an

armoury of epigenetic mechanisms (which do not involve irreversible changes in DNA) to silence genes that impede rapid cell growth. For example, we now know that tumour-suppressor genes can be silenced by mechanisms such as the reversible modification of the regulatory promoters with methyl groups, either on both alleles, or on just one when the other has been deleted⁴.

In addition, there may be not one but several genes with overlapping functions within the LOH region of interest. One example of this might be provided by a region on human chromosome 3 that is frequently lost in lung tumours⁵; pinpointing the critical gene in this section has proved extremely difficult. A strong candidate is infrequently mutated but is often silenced by promoter methylation⁶. To add to this already complicated scenario, 'haploinsufficient' tumour-suppressor genes show loss of function after only one of the two alleles is altered⁷. This raises the spectre of large regions of LOH containing many genes, with one or more key genes being haploinsufficient and therefore hard to find by traditional means, because the corresponding alleles on the undamaged chromosome are not necessarily mutated. Clearly we need to find other approaches.

Functional studies in mice are proving extremely useful here, and lie behind the clever detective work of Li *et al.*³. *Runx1*, *Runx2* and *Runx3* are related mouse genes that encode gene-transcription factors with an impressive list of alternative names and functions⁸; for example, their human counterparts are known to be involved in leukaemias and certain congenital abnormalities. Li *et al.* studied the precise role of *Runx3* by knocking out the gene in mice.

They found that the *Runx3*-deficient

animals died shortly after birth, probably because they couldn't feed properly — the epithelial cells that made up their stomach lining had multiplied excessively. In other words, *Runx3* is involved in keeping cell numbers under control, a typical feature of a tumour suppressor.

A similar effect is seen in mice that lack the transforming growth factor- β 1 (TGF- β 1) protein, which controls cell numbers by inhibiting cell proliferation and inducing cell death. This fact — together with the finding that Smad proteins, which transmit cellular signals coming from TGF- β s, interact directly with the Runx proteins⁹ — led Li *et al.* to investigate the TGF- β 1 signalling pathway in cells from the *Runx3*-deficient mice. The TGF β 1-induced inhibition of proliferation was only modestly affected in *Runx3*-deficient gastric epithelial cells. But the ability of TGF- β 1 to induce cell death was strongly impaired. Moreover, when the authors reintroduced a functional copy of the *Runx3* gene into gastric tumour cells that lacked the protein, tumour growth was inhibited — a strong sign that the authors had identified a gene that suppresses gastric cancers in mice.

Li *et al.* also looked at the expression of the human *RUNX3* gene in patients with gastric cancer. They found that the loss of function of this gene — as a result of deletion of one allele followed by methylation-induced silencing of the other — correlated with the progression of cancer, with all of the eight most advanced tumours studied showing deletions. But the authors' search for the mutational hallmarks of a classical tumour-suppressor gene was unsuccessful. They detected just one *RUNX3* mutation, a 'missense' mutation that alters the encoded protein's structure but may leave some of its functions intact, in 119 tumours. So *RUNX3* is by no means a classical tumour suppressor.

In humans, *RUNX3* is located on the short arm of chromosome 1, in a region designated 1p36 that has long been of interest to cancer researchers. Is *RUNX3* the critical tumour-suppressor gene in this region? No doubt studies of gene expression and promoter methylation, and mutation hunting, will soon provide the solution. In the meantime, there are other questions to answer, not least how *RUNX3* works to keep cell numbers in the stomach under control.

The interaction between the RUNX and SMAD proteins, downstream targets of TGF- β signalling, may be important here. But the answer will not be straightforward. Although SMAD proteins can inhibit cell proliferation, they also induce tumour invasion and metastasis^{10,11}. Moreover, in genetically engineered animal models, TGF- β can inhibit early stages of tumour development but stimulate later tumour progression and metastasis¹⁰. Perhaps the loss of *RUNX3* blocks some of the inhibitory

effects of TGF- β on cell proliferation and death but leaves intact the positive effects of SMADs on malignant progression. In fact, there is evidence that the three *RUNX* genes have multifunctional roles in different tumours and, perhaps, at different stages of cancer^{12,13}.

To muddy the waters still further, although Li *et al.*³ provide convincing evidence that *RUNX3* helps to suppress gastric cancer, it seems that a retrovirus activates the *Runx3* gene in T-cell lymphomas in mice¹⁴. Clearly, much remains to be done to unravel the complexities of this interesting gene family and its roles in cancer.

The bad news about the new wave of tumour-suppressor genes is that they are hard to find. But, unlike the old-style tumour suppressors, they may be good drug targets because they are not mutated, and so could perhaps be reactivated using small molecules that prevent or reverse methylation. Some drugs of this sort are now under development

or in clinical trials¹⁵. Such new-wave drugs may be what is needed to tackle the new generation of tumour suppressors. ■

Allan Balmain is at the UCSF Comprehensive Cancer Center, 2340 Sutter Street, San Francisco, California 94115, USA.

e-mail: abalmain@cc.ucsf.edu

1. Cavenee, W. K. *et al.* *Nature* **305**, 779–784 (1983).
2. Balmain, A. *Nature Rev. Cancer* **1**, 77–82 (2001).
3. Li, Q. L. *et al.* *Cell* **109**, 113–124 (2002).
4. Jones, P. A. & Laird, P. W. *Nature Genet.* **21**, 163–167 (1999).
5. Lerman, M. I. & Minna, J. D. *Cancer Res.* **60**, 6116–6133 (2000).
6. Dammann, R. *et al.* *Nature Genet.* **25**, 315–319 (2000).
7. Philipp-Staheli, J., Payne, S. R. & Kemp, C. J. *Exp. Cell Res.* **264**, 148–168 (2001).
8. Cohen, M. M. *Am. J. Med. Genet.* **104**, 185–188 (2001).
9. Hanai, J. *et al.* *J. Biol. Chem.* **274**, 31577–31582 (1999).
10. Derynck, R., Akhurst, R. & Balmain, A. *Nature Genet.* **29**, 117–129 (2001).
11. Oft, M., Akhurst, R. J. & Balmain, A. *Nature Cell Biol.* (in the press).
12. Osato, M. *et al.* *Blood* **93**, 1817–1824 (1999).
13. Blyth, K. *et al.* *Oncogene* **20**, 295–302 (2001).
14. Stewart, M., MacKay, N., Cameron, E. R. & Neil, J. C. *J. Virol.* **76**, 4364–4369 (2002).
15. Brown, R. & Strathdee, G. *Trends Mol. Med.* **8**, S43–S48 (2002).

Materials science

Cooking up tougher ceramics

D. P. Thompson

Ceramics based on silicon nitride have already proved their worth in terms of engineering applications. A technique that allows greater control over their final structure opens up yet further possibilities.

Microstructural control is at the heart of materials science, especially when a tailored combination of mechanical properties is required. In the field of nitrogen ceramics, where the final microstructure is produced by 'liquid-phase sintering', it is difficult to control the morphology of the final product; the little that can be done is achieved by controlling the starting composition and the processing parameters — principally, time and sintering temperature. Now Shen *et al.*¹ (page 266 of this issue) introduce a technique that opens the way towards significantly greater microstructural control.

Silicon-nitride-based ceramics tend to have the edge over their competitors in engineering applications, thanks to their low density and toughness. When silicon nitride, aluminium oxide and aluminium nitride are reacted together, they form ceramics called 'sialons', after their constituent elements Si, Al, O and N. There are two common forms of sialons, α -sialons and β -sialons, and in their atomic arrangements they are like α - and β -silicon nitride. Since the early 1970s² it has been known that β -sialons readily form with a needle-like (or 'acicular') morphology. The sialon grains may be elongated (with aspect ratios as high as ten), resulting in a tough final product.

Similar acicular morphologies have been

produced for α -sialon ceramics by using unusual starting materials³, an excess of liquid phase or an oxygen-rich starting composition⁴. But 30 years of research have achieved only limited control of the aspect ratio for either α - or β -sialons prepared by conventional sintering procedures.

During liquid-phase sintering of sialons, a chemical force drives the formation of the final phases from the mixture of oxides and nitrides that comprise the starting material. The phases at the final stage of the procedure are essentially the matrix sialon phase (possibly with small amounts of other crystalline phases) and liquid, and these two phases are generally close to thermodynamic equilibrium at the temperature concerned.

Subsequent grain growth occurs by 'Ostwald ripening', whereby the small crystals that are initially present dissolve in the liquid and re-precipitate onto larger crystals, as these are thermodynamically more stable. Generally, the very small grains present in the starting powders will have dissolved rapidly in the early stages of sintering, and, apart from the large acicular β -sialon grains that grow from the silicon nitride needles present in the initial mixture, most of the remaining grains are of similar size. The result is a low driving force for Ostwald ripening, which consequently takes

place slowly, over a period of several hours, even at quite high sintering temperatures.

Moreover, most conventional sintering furnaces reach the final sintering temperature only slowly, and the composition of the liquid phase changes continuously with increasing temperature to remain in equilibrium with the main sialon phase. As a result, there is little additional driving force for grain growth apart from the Ostwald ripening mechanism.

Shen and colleagues¹ use a very rapid sintering technique — 'spark plasma sintering' (SPS) — that enables samples to reach the sintering temperature in less than five minutes. This generates conditions in the sample whereby the composition of the liquid phase is grossly out of equilibrium with that of the matrix sialon grains, thereby creating a strong chemical driving force for the liquid and matrix to equilibrate. The resulting transport of material provides an alternative mechanism for grain growth — the so-called dynamic Ostwald ripening effect. For both α - and β -sialon phases, this results in elongated, needle-shaped grains; the increase in length of the needles after such short sintering times is quite remarkable.

The results also show that the process has a minimum temperature threshold. This might be related to the properties of the liquid — for example, unless the temperature is high enough, the liquid may be too viscous to allow the molecules to diffuse easily. Alternatively, high temperatures may be required to break the covalent bonds in the silicon and aluminium nitride powders so that they dissolve in the liquid.

Shen *et al.* carried out their work in a research-scale SPS furnace, and scaling up the process to allow large-scale production of sialon components with a controlled high-aspect-ratio acicular grain morphology may seem a long way off. Nevertheless, the dynamic Ostwald ripening principle seems applicable to all nitrogen ceramics, regardless of the matrix form (α or β) or the sintering additive (another way of modifying the liquid's properties). The door has been opened and, without doubt, further use of SPS will allow far more detailed study of the grain growth mechanism in nitrogen ceramics. As a result, the principles governing the development of tough, needle-shaped microstructures will be determined with greater precision, and the prospects look good that a new range of high-toughness sialon ceramics will emerge. ■

D. P. Thompson is in the Materials Division, MMME Department, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, NE1 7RU, UK.

e-mail: D.P.Thompson@ncl.ac.uk

1. Shen, Z., Zhao, Z., Peng, H. & Nygren, M. *Nature* **417**, 266–269 (2002).
2. Drew, P. & Lewis, M. H. *J. Mater. Sci.* **9**, 261–269 (1974).
3. Chen, I.-W. & Rosenflanz, A. *Nature* **389**, 701–704 (1997).
4. Zhao, H., Swensen, S. P. & Cheng, Y.-B. *J. Eur. Ceram. Soc.* **18**, 1053–1057 (1997).

How to be desensitized

Christian Rosenmund and Michael Mansour

The various functional states of glutamate receptors control much of the brain's neuronal activity. Our understanding of how one of those states — desensitization — occurs has taken a leap forward.

Neurons communicate with each other through synapses, which operate in the following way. On one side of the synapse, the incoming neuron converts electrical activity, encoded in action potentials, into a chemical signal by releasing a neurotransmitter. The membrane of the receiving, postsynaptic, neuron is packed with ion channels, which transduce neurotransmitter binding into complex conformational changes that induce channel opening and closing, a process known as gating. Channel gating translates the concentration profile of neurotransmitter in the synapse into a defined time course of ion flow across the postsynaptic neuronal membrane, changing the membrane potential accordingly and, usually, resulting in onward transmission of the presynaptic action potential. The most important class of neurotransmitter-gated ion channels conveying excitatory signals in the brain are the AMPA-type glutamate receptors. On page 245 of this issue, Sun *et al.*¹ provide us with exciting snapshots of the gating process involved.

Like many other ion channels, AMPA receptors are doughnut-shaped structures, formed by the assembly of different members of the family of AMPA-receptor subunits², that span the insulating neuronal

membrane, with an ion-conducting water-filled pore down the centre. The pore is controlled by a gate, which 'opens' to permit ion flow only when the neurotransmitter — glutamate — binds to the receptor. However, AMPA receptors share with other ion channels a property known as 'desensitization', which limits the effective ion flow to the few milliseconds following binding of glutamate. This is because the channel tends to close even though neurotransmitter is still bound to it. This process has a considerable influence on how AMPA receptors encode information, as the degree of desensitization is a major factor in shaping the time course of postsynaptic responses³. Moreover, the degree of desensitization of the AMPA receptor is under exquisite control in various ways, such as regulated gene expression, RNA editing, alternative splicing of receptor-encoding RNAs, and the assembly of receptors from different subunits². Despite the great importance of desensitization in brain function, however, little is known about how AMPA-receptor subunits coordinate gate opening and desensitization.

The extracellular glutamate-binding portion of each subunit comprises two protein domains. Earlier studies⁴ provided the crystal structure of these domains, which can

best be described as resembling a clam shell. On binding glutamate, the cleft between the two shells (the two domains) narrows by the movement of the second domain towards the first⁵. This may directly trigger the opening of the gate, as amino-acid residues that link domain 2 to the plasma membrane are pulled away from the pore-forming region of the receptor. Because AMPA receptors are tetramers, and because all subunits are able to bind glutamate, they may unite in a step-wise fashion in the attempt to open the gate⁶. An unusual structural feature is that individual glutamate-binding pockets of AMPA receptors pair up to form dimers⁵. The dimer interface, which is formed by a twofold symmetric assembly through an interaction at the domain 1 interfaces, is packed with residues that change receptor desensitization dramatically if structurally modified³, indicating that the dimer interface may be involved in desensitization.

Sun *et al.*¹ have used a combination of techniques in their study, and the results, taken together, are a major step forward in our understanding of how the AMPA receptor gates. First, they show that the stability — or rather lack of stability — of the dimer complex seems to be crucial for the desensitization process. They took the water-soluble glutamate-binding portions of the receptors (which they otherwise used to obtain the crystals) and measured their tendency to dimerize by sedimentation assay. When they compared the affinities with functional measurements of desensitization in the intact receptor, they found that a mutation that blocks desensitization⁷ increases the likelihood of dimerization of individual subunits by a factor of a hundred thousand. Similarly, dimerization of normal desensitizing receptors is greatly increased by the potent blocker of desensitization, cyclothiazide. Conversely, a mutation that accelerates desensitization reduces dimer affinity to a degree that it almost escapes detection. Comparison of dimer affinity and desensitization revealed an almost perfect inverse correlation between the two.

Next, with the aim of revealing the structural basis of the dimerization process, Sun *et al.* compared the crystal structures of mutant receptors that had dramatically different desensitization properties. The increased dimer affinity of the non-desensitizing mutant can be nicely explained by the formation of a strong hydrophobic pocket in the dimer interface. An equivalent effect can be seen in crystals of normal — wild-type — desensitizing receptors in the presence of cyclothiazide.

So what does the desensitization process look like? The authors' view of events is shown in Fig. 5 of the paper (page 251) and in Fig. 1 here. Activation of the receptor by glutamate, and subsequent movement of domain 2, produces tension on the

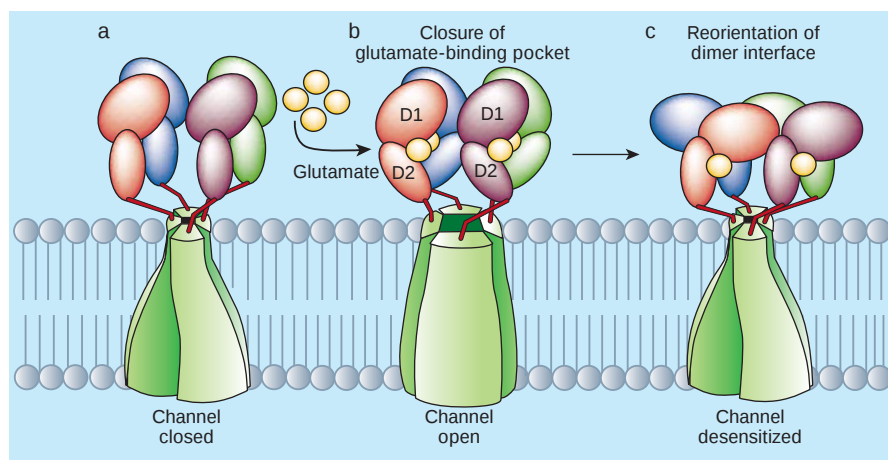


Figure 1 Activation and desensitization of AMPA-type glutamate receptors. **a**, The glutamate-binding domains are organized as a 'dimer of dimers', with the red and blue glutamate-binding cores forming one dimer and the purple and green subunits forming the other. The two dimers sit side-by-side so that the inter-dimer local twofold axis is in the centre of the four subunits and coincides with the fourfold axis of the ion channel. **b**, In the scheme of events proposed by Sun *et al.*¹, binding of glutamate (yellow particles) induces domain closure at the glutamate-binding core: movement of domain 2 (D2) towards domain 1 (D1) causes conformational changes at the channel gate and opens the channel. **c**, Reorientation of the dimer interface decouples glutamate-binding-induced domain closure from the channel gate, and leads to receptor desensitization. (Modified from a graphic provided by Y. Sun.)

connecting portion between the extracellular binding pocket and the receptor's transmembrane domains. This leads to conformational changes within the membrane, forcing the gate to open. But in consequence there is an equivalent strain on the glutamate-binding pocket, facilitating an ensuing conformational rearrangement to escape tension. This causes rearrangement at the dimer interface and movement of the receptor into the desensitized state.

In receptors that do not desensitize, the strengthened dimer interface prevents rearrangement within the ligand-binding pocket. The linker between binding pocket and gate remains under tension and leaves the pore open as long as glutamate is bound. It is therefore no surprise that the glutamate-binding domain of the wild-type receptor and that of the non-desensitizing mutant both crystallize into the same 'open-channel' state, because the link, and therefore the tension force, from the membrane is missing.

Finally, Sun *et al.* surprise us with crystal structures of a mutant receptor that was designed to disrupt the dimer interface. The accelerated desensitization properties of the intact mutant receptor nicely confirm the importance of dimer affinity in desensitization. As expected, the structure does not dimerize at the usual interface. But it does reveal another, previously unknown, subunit interaction — hinting at a possible way in which pairs of dimers can assemble into a tetramer. In this interaction, domain 1 connects with a different region (lateral to that involved in dimerization) of domain 2 of a neighbouring receptor. If this interaction indeed takes place in the assembled receptor complex, the tetrameric structure represents a pair of dimers that are shifted parallel to one another.

This would be an unprecedented form of channel subunit assembly — one in which the dimers facing each other follow a different symmetry from the assumed (though not proven) fourfold symmetry of the receptor in the transmembrane channel pore. This arrangement may, however, provide a positional clue to how different isoforms of AMPA-receptor subunits assemble into a complex *in vivo*⁸, as chemically equivalent subunits may prefer to assume positions of equivalent symmetry, namely diagonally across from each other. It is yet not clear how these arrangements would affect receptor gating. But it is clear that glutamate receptors are likely to remain a rich source of surprises. ■

Christian Rosenmund and Michael Mansour are in the Department of Membrane Biophysics, Max-Planck-Institute for Biophysical Chemistry, D-37077 Göttingen, Germany.
e-mail: crosenm@gwdg.de

4. Armstrong, N., Sun, Y., Chen, G. Q. & Gouaux, E. *Nature* **395**, 913–917 (1998).
5. Armstrong, N. & Gouaux, E. *Neuron* **28**, 165–181 (2000).
6. Rosenmund, C., Stern-Bach, Y. & Stevens, C. F. *Science* **280**, 1596–1599 (1998).
7. Stern-Bach, Y., Russo, S., Neuman, M. & Rosenmund, C. *Neuron* **21**, 907–918 (1998).
8. Mansour, M., Nagarajan, N., Nehring, R. B., Clements, J. D. & Rosenmund, C. *Neuron* **32**, 841–853 (2001).

Developmental biology

Signalling polarity

Randall T. Moon and Kavita Shah

Cell signalling pathways triggered by Wnt proteins control gene expression and cell behaviour, especially during development. Work on frogs reveals another specific role for these proteins, and the signalling involved.

Living organisms tend to be highly economical, using the same proteins or cell-to-cell signalling processes in different contexts and for different purposes. Take, for example, the family of proteins known as NFATs. These are transcription factors — they regulate gene expression — and they have traditionally been implicated in controlling genes involved in immunity. The events leading up to such regulation include an increase in the levels of calcium ions in a cell; the NFAT transcription factors then move into the nucleus and form a complex with relevant regions of DNA (and, in some cases, with other transcription factors¹). Several different signalling pathways are known to activate NFAT by changing calcium levels, and to this list can now be added one of the pathways that is triggered by Wnt proteins, as Saneyoshi and colleagues² describe on page 295 of this issue. Rather than being involved in immunity, however, these events are key to early vertebrate development.

Members of the Wnt protein family are secreted from cells and picked up by receptor

proteins on the surface of cells, triggering intracellular signalling pathways. These pathways regulate cell proliferation, death, fate and behaviour in contexts ranging from early embryonic development to colorectal cancer. The best-understood Wnt pathway is the 'canonical' one: here, Wnt-induced signalling suppresses degradation of the β -catenin protein, enabling it to accumulate in the nucleus³. Nuclear β -catenin then binds to particular transcription factors (not the NFATs) and thereby modulates gene expression.

Over the past few years it has become apparent that certain Wnt proteins and their receptors (proteins of the Frizzled family) can also activate 'non-canonical' pathways. Although the details of the canonical Wnt/Frizzled pathway are firmly supported by genetic, biochemical and cell-biological experiments in many species, the non-canonical pathways are less well understood. Nonetheless, it is known that a non-canonical 'planar cell polarity' pathway is used in fruitflies to orientate cells⁴.

There also seems to be more than one

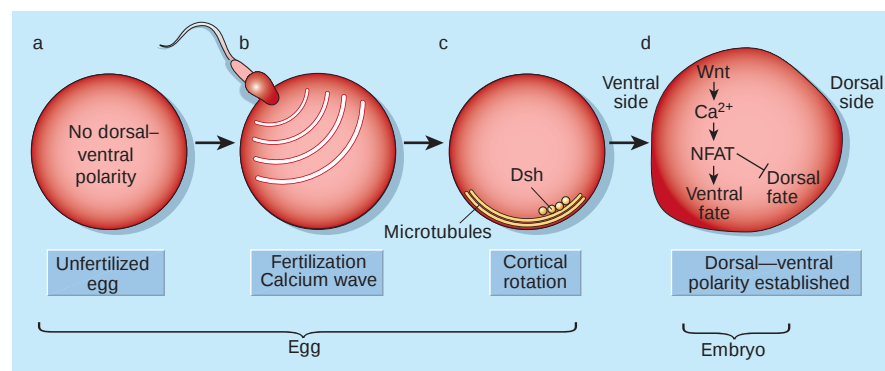


Figure 1 Opposing signalling pathways, triggered by different members of the Wnt protein family, may help determine the polarity of *Xenopus* embryos. **a**, Before fertilization, there is no evident distinction between the dorsal and ventral sides of the egg. **b**, Fertilization by a sperm anywhere in the upper half of the egg initiates a wave of Ca²⁺ ions, which originate in what will become the animal's ventral side. **c**, During the first hour after fertilization, the cortex (the plasma membrane and associated structures and molecules) rotates by about 30° relative to the inner cytoplasm. The Dishevelled (Dsh) protein, a downstream target of 'canonical' Wnt pathways, moves along filamentous tracks (microtubules) to the future dorsal side, where it will ultimately result in stabilization of the β -catenin protein in cell nuclei, leading to transcription of relevant genes. **d**, Saneyoshi *et al.*² predict that sometime between fertilization and the establishment of dorsal-ventral polarity, NFAT is activated in response to Wnt-triggered rises in Ca²⁺ levels in cells on the future ventral side. This promotes ventral cell fates, and antagonizes the signal that promotes dorsal cell fates.

1. Sun, Y. *et al.* *Nature* **417**, 245–253 (2002).
2. Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. *Pharmacol. Rev.* **51**, 7–61 (1999).
3. Jones, M. V. & Westbrook, G. L. *Trends Neurosci.* **19**, 96–101 (1996).

Wnt/Frizzled pathway at work in vertebrate embryos. Early studies of the South African clawed frog, *Xenopus laevis*, established that the overexpression of some Wnt proteins stabilizes β -catenin and results in twinned embryos, complete with two heads. Yet overexpression of other Wnt proteins perturbs cell movements during gastrulation³ — a crucial stage of development that leads to formation of the three main tissue layers: endoderm, mesoderm and ectoderm. These different effects hinted that there is more than one Wnt pathway in vertebrates.

The idea was substantiated by further work in zebrafish and *Xenopus* embryos, which suggested that the Wnts that perturb cell movements activate a non-canonical Wnt/Frizzled pathway³ involving increases in intracellular Ca^{2+} levels and the consequent activation of the enzymes protein kinase C (PKC) and Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII)⁵. But questions remain, including the relationship between the non-canonical pathways in fruitflies compared with vertebrates, whether the Wnt/ Ca^{2+} pathway works in any context other than early embryos, and whether it modulates gene expression or only gastrulation movements.

Saneyoshi *et al.*² enter this arena of uncertainty by proposing that if signalling from certain Wnt proteins leads to rises in intracellular Ca^{2+} levels, it may also activate calcineurin, a Ca^{2+} /calmodulin-dependent protein phosphatase, leading to the removal of phosphate groups from NFAT and hence its accumulation in the nucleus. And indeed, they find that Wnt5A and Frizzled-2 — the same Wnt–receptor combination reported to lead to intracellular Ca^{2+} increases in zebrafish and to activate CaMKII and PKC in *Xenopus* — strongly induces the movement of NFAT into the nucleus in cells of *Xenopus* embryos.

The authors also observe that overexpression of a constitutively active NFAT perturbs the cell movements that occur during gastrulation in *Xenopus*, and encourages cells to take on characteristics typical of the belly (ventral) rather than the dorsal part of the embryo. Finally, this hyperactive NFAT antagonizes the duplication of the dorsal axis that can be induced by the canonical Wnt pathway. Similar effects have been seen³ with overexpression of Wnt5A. So Saneyoshi *et al.*'s findings support the existence of the Wnt/ Ca^{2+} pathway in *Xenopus*, and identify NFAT as a likely downstream target of this pathway.

What, then, is the most plausible role of normal Wnt/ Ca^{2+} signalling and NFAT in early frog embryos? Saneyoshi *et al.*'s data lend much weight to the previous suggestion⁵ that this non-canonical pathway helps to determine a ventral fate for cells and opposes the canonical pathway, which promotes dorsal cell fates in early embryos (Fig. 1, previous page). In support of this, Saneyoshi *et al.* show that expressing a

presumed inhibitor of NFAT in frog embryos results in a new dorsal axis, as would be predicted if the NFAT had been promoting ventral cell fates.

So the new data are a tantalizing hint that Wnt/ Ca^{2+} signalling promotes ventral cell fates by activating NFAT. But there are questions to be answered before we can be sure. First, does NFAT actually activate gene transcription in response to a Wnt/ Ca^{2+} signal? Saneyoshi *et al.* show that, in the presence of one of its cooperative partners — the transcription factor AP-1 — a constitutively active *Xenopus* NFAT activates transcription of a 'reporter' gene inserted into *Xenopus* embryos. But it is not yet certain whether normal NFAT would do so in response to Wnt/ Ca^{2+} signalling, or whether Wnt5A leads to the transcription of NFAT-responsive genes, as would be predicted. There is evidence that Wnt5A regulates AP-1-dependent transcription as well as Ca^{2+} -mediated signalling⁶. So the Wnt/ Ca^{2+} gene targets may encompass those regulated by AP-1 alone, or those regulated by complexes of AP-1 and NFAT.

Second, Saneyoshi *et al.*'s data predict that more NFAT proteins will be found in cell nuclei on the future ventral side of *Xenopus* embryos than on the future dorsal side. If that is the case, does this nuclear accumulation depend on Wnt signals? The activity of CaMKII is increased on the future ventral side in a Wnt-dependent manner⁵, but it remains to be seen whether the nuclear accumulation of NFAT is likewise increased. Third, we need to test the proposed role of NFAT by knocking out its function, using genetic or molecular means.

Finally, it will be interesting to see whether the Wnt5A-driven movement of NFAT to the nucleus is unique to early vertebrate development, where all characterization of the Wnt/ Ca^{2+} pathway has been conducted to date. Unpublished work from Chris Hughes (Univ. California, Irvine) suggests the contrary: that this Wnt protein also induces NFAT to accumulate in the nuclei of the immune system's T cells. Whatever the answers, it is clear that it takes many Wnt pathways to make a frog — and, by extension, any vertebrate. ■

Randall T. Moon and Kavita Shah are at the Howard Hughes Medical Institute and Department of Pharmacology, University of Washington School of Medicine, Seattle, Washington 98195, USA.
e-mails: rtmooon@u.washington.edu
kvshah@u.washington.edu

- Macian, F., Lopez-Rodriguez, C. & Rao, A. *Oncogene* **20**, 2476–2489 (2001).
- Saneyoshi, T., Kume, S., Amasaki, Y. & Mikoshiba, K. *Nature* **417**, 295–299 (2002).
- Kuehl, M., Sheldahl, L. C., Park, M., Miller, J. R. & Moon, R. T. *Trends Genet.* **16**, 279–283 (2000).
- Axelrod, J. D. & McNeill, H. *Sci. World J.* (in the press).
- Kuehl, M., Sheldahl, L. C., Malbon, C. C. & Moon, R. T. *J. Biol. Chem.* **275**, 12701–12711 (2000).
- Yamanaka, H. *et al.* *EMBO Rep.* **3**, 1–7 (2002).

Daedalus

Augmented eggs

The size of an egg is one of evolution's battlegrounds. The chick wants the biggest possible egg, to aid its development; the mother wants the smallest possible egg, to help her to lay it. Neither party gets its ideal. Daedalus now has a compromise. He wants the egg to expand after being laid.

Sadly, all birds' eggs have a rigid carbonate shell. So DREADCO biologists are breaking fertilized eggs with great care. They are opening them into larger glass or plastic cavities, filled with distilled water or nutritive solution and sealed to the unbroken section of shell with silicone resin. The chick within will find plenty of room to develop further, before having to break its way out. Mother birds may not want to sit on and hatch augmented eggs, although some females appreciate outsize ones. Design problems will loom large. It will probably be easier to hatch augmented eggs in an incubator.

In normal conditions, most birds hatch fairly simple-minded offspring — Daedalus recalls Konrad Lorenz's goslings, which followed him because he pretended to be their mother. But a chick that develops in an expanded egg would take longer in development, and could be far shrewder than average. Even more cunning, imagine two eggs sealed together by a tube the diameter of one of them, the extra space filled with nutrient solution. Would one chick be aware of the other? Would they develop as a clever pair, tackling problems neither could work out alone? Might they even have some subtle avian empathy? A whole new type of bird might result from this research. It could emerge from the augmented egg with much enhanced mental or physical powers.

Daedalus is not sure whether to try the idea on domestic fowl. They are selected for stupid tractability; bright ones could be very troublesome. But ducks seem to have a lot of enterprise already. They could be ideal. The new improved chicks might swim or fly better, or be stronger and more decisive. They could come to dominate their fellows. Even so, an augmented duck that came to dominate a flock of them would be unable to hand on its abilities to the next generation. So Daedalus's egg augmenters are concentrating on birds that have recently lost the power of flight, such as penguins, kakapos and certain rheas. Is the ability to fly still inherent in a chick? If so, could it be recovered by egg augmentation?

David Jones

An ancient sensory organ in crocodilians

Waiting alligators can detect silent ripples in the water even in total darkness.

Crocodilians hunt at night, waiting half-submerged for land-bound prey to disturb the water surface. Here I show that crocodilians have specialized sensory organs on their faces that can detect small disruptions in the surface of the surrounding water, and which are linked to a dedicated, hypertrophied nerve system. Such 'dome' pressure receptors are also evident in fossils from the Jurassic period, indicating that these semi-aquatic predators solved the problem of combining armour with tactile sensitivity many millions of years ago.

During behavioural trials, half submerged alligators (*Alligator mississippiensis*) orientated themselves to single water droplets in complete darkness without the use of audition (Fig. 1a, b). This orientating behaviour was robust and consisted of either head-turning or a full body lunge towards the source. It occurred only when the alligators' faces were located on the air–water interface, and not when the animals were

completely submerged or had their heads entirely out of the water.

The faces of alligators are covered with small, pigmented domes (Fig. 2), the distribution of which resembles a beard, although they are also present inside the mouth. To test whether these organs are involved in head-orientating behaviour, I covered them with a plastic elastomer. The orientating behaviour was abolished completely.

Such organs were previously named integumentary sensory organs¹ and were used to segregate the alligatorid and crocodilian families². However, their function, innervation and phylogenetic history were not known. On the basis of my observations, I prefer a functionally descriptive name for these organs: dome pressure receptors (DPRs).

DPRs are round, dome-like structures which lack pores or protruding hairs (Fig. 1c, inset). The epidermis is 40% thinner immediately above the DPRs, whereas the

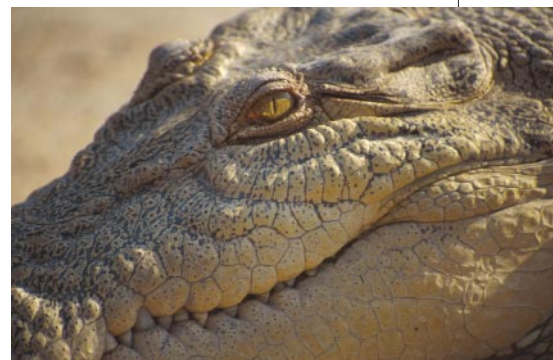


Figure 2 Stubble means trouble: an alligator's face, showing the distribution of the blue dome pressure receptors.

keratin layer is 60% thinner and more compact. A dermal fold below each organ contains a highly branched nerve bundle. Tract-tracing reveals secondary and tertiary branching directly under the epidermis, and trigeminal innervation. It is not surprising that the trigeminal nerve innervates DPRs, for the trigeminal pathway innervates many specialized organs in vertebrates, such as the electrosensory organs in the platypus³ and infrared detectors in snakes⁴. The trigeminal nerve, which leads to a hypertrophied brain-stem nucleus, is the largest of all the cranial nerves in the alligator.

Recordings from the trigeminal ganglion indicated that DPRs are sensitive to surface waves, created by using a surface-wave generator or a piezoelectric 'poker'. As it proved difficult to manipulate the magnitude and frequency of the same droplet stimulus used in behavioural trials, I used a surface-wave generator to measure responses. The generator consisted of a sinusoidal-wave generator attached to a rod and connected to a speaker. Surface waves evoked a single spike phase locked to the stimulus. Increasing the wave amplitude caused an increased probability of spike firing (Fig. 1d).

To stimulate individual receptors and the skin between them, I used a blunt, closed capillary tube attached to a piezoelectric wafer as a secondary stimulus. Under continuous pressure, DPR responses adapted quickly. There was no response when the skin between DPRs was subjected to the same pressure that evoked a response in the DPRs. This supports the idea that DPRs are responsible for detecting the pressure waves associated with disruptions to the air–water interface. Although the mechanism of transduction is unknown, it is most likely that DPRs are sensitive to pressure differences and not to particle motion, although responses to shear cannot be ruled out.

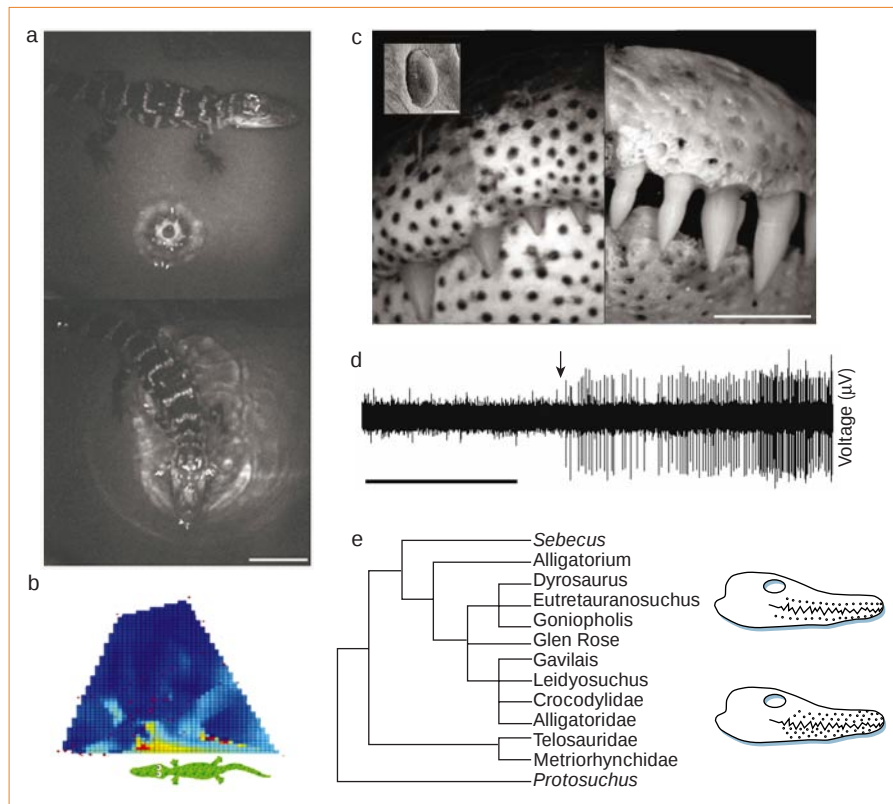


Figure 1 Dome pressure receptors (DPRs) in crocodilians. **a**, Infrared images of the responses of an alligator to water droplets. Scale bar, 10 cm. **b**, Summary of the responses of one alligator. Cartoon represents the position of the animal; colour represents an error index as follows: blue, <10% error; red, 150% error. **c**, Left, DPRs can be seen as black dots; right, corresponding foramina. Scale bar, 3 cm. Inset, scanning electron micrograph of an individual DPR; scale bar, 100 μ m. **d**, Physiological trace from a trigeminal ganglion showing the responses of DPRs to increases in water pressure (left to right). Arrow indicates the threshold of the response. Scale bar, 10 s. **e**, Cladogram of crocodyliforms^{5,6} showing the presence and absence of foramina. Forms in which the pattern of foramina is absent (top drawing) are shown in italics.

The phylogenetic history of these receptors can be determined because the DPR innervation is evident in the mandibular and maxillary bones as a hexagonal pattern of foramina for the branches of the trigeminal nerve. I examined skulls on the premise that this unique pattern of foramina corresponds to the presence of DPRs (Fig. 1c, right). All living crocodilians examined had foramina and DPRs.

The skulls of members of the four living lizard families (Iguania, Gekkota, Scincophora and Anguimorpha) did not show this pattern of foramina, even though these animals also possess cutaneous specializations for tactile sensitivity. Unlike crocodilians, all lizards have a linear pattern of foramina on the top and bottom of the tooth line and none shows a 'beehive' pattern, including the marine iguana (*Amblyrhynchus cristatus*), which inhabits both land and water environments.

I also examined extinct crocodilian forms. Only animals that are thought to have had a semi-aquatic lifestyle show the same pattern of foramina as living crocodilians. Extinct, fully terrestrial crocodilians (such as *Sebecus*) do not show this pattern of foramina and the most ancient crocodilian specimen examined (*Protosuchus*) shows the lizard-like linear pattern. DPRs thus seem to

have emerged in the Early Jurassic (Fig. 1e).

Crocodylians are an ancient monophyletic group that has existed since before the time of the dinosaurs. Behavioural, anatomical, physiological and palaeontological evidence indicates that these predatory animals have evolved a unique sensory organ to mediate orientation towards disruption of the water surface. DPRs on crocodilian skin are stimulated by surface waves and covering them abolishes the animal's orientating behaviour. Osteological markings of DPRs appear on fossil material in the form of foramina, which suggests that this receptor is a unique, conserved feature of semi-aquatic crocodilians.

Daphne Soares

Department of Biology, University of Maryland,
College Park, Maryland 20742, USA
e-mail: daph@wam.umd.edu

1. von Düring, M. Z. *Entwicklungsgesch.* **143**, 81–94 (1973).
2. Jackson, K., Butler, D. G. & Youson, J. J. *Morphol.* **229**, 315–324 (1996).
3. Gregory, J. E., Iggo, A., McIntyre, A. K. & Proske, U. *Nature* **326**, 386–387 (1987).
4. Molenaar, G. J. *Brain Res.* **78**, 340–344 (1974).
5. Benton, M. J. & Clark, J. M. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 295–338 (Clarendon, Oxford, 1988).
6. Clark, J. M. in *In the Shadow of the Dinosaurs: Early Mesozoic Tetrapods* (eds Fraser, N. & Sues, H.-D.) 84–97 (Cambridge Univ. Press, 1994).

Competing financial interests: declared none.

Acidic deposition

Decline in mobilization of toxic aluminium

The mobilization of aluminium from acidic forest soils is arguably the most ecologically important consequence of acid deposition in the environment because of its adverse effects on soils, forest vegetation and surface water^{1–3}. Here we show that there has been a significant decline in the concentrations of aluminium species in soil solutions at medium-to-high elevations in a northern hardwood forest in the United States in response to decreasing acidic deposition. Streamwater aluminium concentrations have also fallen and, if this rate of recovery persists, will within 10 years no longer pose a threat to fish.

Atmospheric deposition of sulphur has been declining over the past 30 years in response to controls on sulphur dioxide (SO₂) emissions in many parts of Europe and North America, resulting in reduced concentrations of sulphate (SO₄^{2−}) in drainage and surface waters^{4–6}. Although limited recovery in surface-water pH and acid-neutralizing capacity have been reported for some regions, this has not occurred in the northeastern United States, where the atmospheric input of base cations has declined and where soil pools

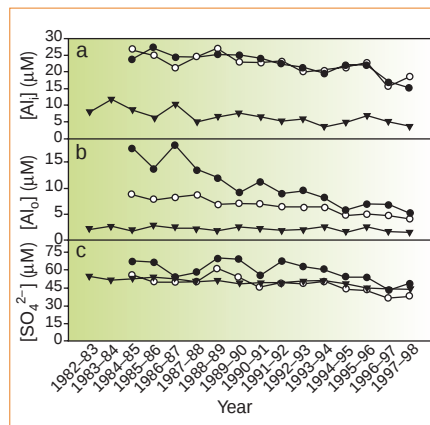


Figure 1 Long-term declines in sulphate concentrations mitigate the mobilization of aluminium in soil solutions and stream water in the Hubbard Brook Experimental Forest, New Hampshire, USA. **a–c**, Annual volume-weighted concentrations of: **a**, inorganic monomeric aluminium (Al_i); **b**, organic monomeric aluminium (Al_o); and **c**, sulphate (SO₄^{2−}) in mineral soil solutions at 750 m (filled circles) and 730 m (open circles), and in stream water (triangles).

of these ions are low^{4–6} and have been depleted by acid deposition⁷. Critically, acid deposition causes an increase in aluminium concentrations, but resulting trends in soil and surface waters have not been described.

We examined long-term trends in soil solutions (1984–98) and stream water (1982–98) from reference watershed 6 in the Hubbard Brook Experimental Forest

(HBEF) in New Hampshire, USA. The primary source of soil aluminium at this site is weathering of aluminium-rich minerals in tills and bedrock; over the years, the input of strong-acid anions from atmospheric deposition has increased the mobilization of aluminium ions in soil solutions and surface waters^{8–10}.

We observed a significant decline in the concentrations of both inorganic and organic monomeric aluminium (Al_i and Al_o, respectively) in solutions draining from mineral soils at medium-to-high-elevation sites (730 and 750 m above sea level; Fig. 1). Concentrations of Al_i declined from 23.7 to 15.2 μM at 750 m, and from 26.6 to 18.6 μM at the 730-m site, over the period of study (annual decline, 0.60 and 0.42 μM yr^{−1}, respectively; *P* < 0.001). Concentrations of the potentially less toxic Al_o also declined significantly (0.9 μM yr^{−1} at 750 m; 0.4 μM yr^{−1} at 730 m; *P* < 0.001).

These decreases in aluminium concentration in soil solutions seem to be mediated by declining SO₄^{2−} concentrations (Fig. 1). There was no change in aluminium concentration in soil solutions draining from surface organic soils, and there was also no long-term trend evident in solutions draining from soils at a lower elevation (600 m).

During 1982–98, annual volume-weighted concentrations of Al_i in stream water also decreased (0.2 μM yr^{−1}; *P* < 0.001), and will fall to below the toxic threshold for fish³ within about 10 years if current rates of decline are maintained. However, fish survival probably responds to peak conditions³ rather than to average conditions. Aquatic biota may still be at risk from intermittently high concentrations of Al_i in stream water; concentrations in excess of 8 μM were recorded at HBEF during snow-melt events in 2001.

Despite the marked decrease in SO₄^{2−} concentrations in stream water (1.4 micro-equivalents per litre per year; *P* < 0.001), we observed only a modest increase in acid-neutralizing capacity (0.8 μEq l^{−1} yr^{−1}; *P* < 0.01) and no change in stream pH. Not only has the depletion of soil base-cation pools limited the rate of stream recovery from acidic deposition^{4,7}, but the hydroxylation of aluminium also seems to have been critical in buffering waters against increases in pH and acid-neutralizing capacity.

Sheila M. Palmer, Charles T. Driscoll

Department of Civil and Environmental
Engineering, Syracuse University, 220 Hinds Hall,
Syracuse, New York 13244, USA
e-mail: shpalmer@syr.edu

1. Cronan, C. S. & Grigal, D. F. *J. Environ. Qual.* **24**, 209–226 (1995).
2. Driscoll, C. T. & Postek, K. M. in *The Environmental Chemistry of Aluminium* (ed. Sposito, G.) 363–418 (CRC, Boca Raton, Florida, 1996).
3. Bulger, A. J., Lien, L., Cosby, J. & Henriksen, A. *Can. J. Fish. Aquat. Sci.* **50**, 575–585 (1993).

4. Likens, G. E., Driscoll, C. T. & Buso, D. C. *Science* **272**, 244–246 (1996).
5. Driscoll, C. T. et al. *BioScience* **53**, 180–198 (2001).
6. Stoddard, J. L. et al. *Nature* **401**, 575–578 (1999).
7. Likens, G. E. et al. *Biogeochemistry* **41**, 89–173 (1998).
8. Johnson, N. M., Driscoll, C. T., Eaton, J. S., Likens, G. E. & McDowell, W. J. *Geochim. Cosmochim. Acta* **45**, 1421–1437 (1981).
9. Driscoll, C. T., van Breemen, N. & Mulder, J. *Soil Sci. Soc. Am. J.* **49**, 437–444 (1985).
10. Lawrence, G. B., Fuller, R. D. & Driscoll, C. T. *Biogeochemistry* **2**, 115–135 (1986).

Competing financial interests: declared none.

Caenorhabditis elegans

Plague bacteria biofilm blocks food intake

Bubonic plague is transmitted to mammals, including humans, by the bites of fleas whose digestive tracts are blocked by a mass of the bacterium *Yersinia pestis*¹. In these fleas, the plague-causing bacteria are surrounded by an extracellular matrix of unknown composition², and the blockage depends on a group of bacterial genes known as the *hmsHFRS* operon³. Here we show that *Y. pestis* creates an *hmsHFRS*-dependent extracellular biofilm to inhibit feeding by the nematode *Caenorhabditis elegans*. Our results suggest that feeding obstruction in fleas is a biofilm-mediated

process and that biofilms may be a bacterial defence against predation by invertebrates.

A biofilm is a population of adherent bacteria enclosed by a matrix⁴. When *C. elegans* is exposed to *Y. pestis* or to the closely related bacterium *Yersinia pseudotuberculosis*, biofilms become visible on the worms' heads within 1 h. The biofilms increase in size with continuing exposure, eventually covering the mouth completely (Fig. 1a). Bacterial cells rarely come into direct contact with the worms — they are not found at all in interior tissues (Fig. 1a, b) and there is no other evidence of infection.

C. elegans and many other invertebrates eat bacteria, and microbial synthesis of a biofilm that inhibits predatory feeding therefore has obvious survival value. To demonstrate that the biofilm produced by *Yersinia* prevents feeding by *C. elegans*, we allowed biofilms to accumulate on *phm-2* mutant nematodes⁵, which pass many unlysed bacteria through the pharynx when feeding. We placed the worms on lawns of *Escherichia coli* expressing green fluorescent protein (GFP) and monitored the presence of GFP in the gut, which indicated an ability to feed.

The intestines of 99.0% ($n=238$) of unexposed controls (Fig. 1c) were positive for GFP, whereas only 2.9% ($n=171$) of worms exposed to *Y. pseudotuberculosis* (Fig. 1d) and 46.1% ($n=107$) of those exposed to *Y. pestis* were positive. The lesser

inhibition by *Y. pestis* was due to the organism's nutritional defects⁶ — when grown on a richer medium, it allowed only 7.8% ($n=166$) of worms to feed.

C. elegans grows through four larval stages to adulthood, but animals that are deprived of food fail to develop normally. Starting with eggs of uniform age, 99.4% of animals grown on an *E. coli* control strain developed to the fourth larval (L4) stage within 2 days. For worms exposed to *Y. pestis* and *Y. pseudotuberculosis*, the proportions were 6.9% and 43.9%, respectively. With *Y. pestis* grown on richer medium, only 2.4% of nematodes reached the L4 stage.

Random screening produced a transposon-insertion mutant of *Y. pseudotuberculosis* that allowed normal nematode growth and failed to produce a biofilm. The mutation was in *hmsT*, which was previously identified as a positive regulator of the *hmsHFRS* operon in *Y. pestis*⁷. We created *Y. pestis* and *Y. pseudotuberculosis* strains with mutations in *hmsHFRS* and found that these bacteria failed both to produce biofilms and to inhibit nematode growth.

Homologues of HmsF and HmsR are required for the synthesis of polysaccharide components of biofilms by *Staphylococcus epidermidis*^{8,9}. HmsR is 39% identical to staphylococcal IcaA, a glycosyltransferase that catalyses the assembly of large polymers of *N*-acetyl glucosamine residues⁸. HmsF is 23% identical to IcaB, a predicted staphylococcal polysaccharide deacetylase.

In fleas, plague bacteria blocking the digestive tract are surrounded by an electron-dense matrix, the composition of which has not been described². Because flea blockage and biofilm formation on nematodes both require HmsHFRS proteins, our results suggest that the matrix in fleas also includes a bacterially synthesized biofilm. Because biofilms are highly resistant to flushing with fluids, the presence of a biofilm may help to explain why *Y. pestis* is not dislodged from the flea gut by the pumping and threshing action of the insect's digestive system¹⁰.

Genetic analysis of both sides of the bacteria–nematode interaction is feasible, as *C. elegans* mutants that are resistant to biofilm formation can readily be obtained by chemical mutagenesis (C.D., unpublished results). The *Yersinia*–*C. elegans* model therefore offers a potentially useful experimental system in which to investigate biofilm-mediated interactions between bacteria and invertebrates.

Creg Darby, Jennifer W. Hsu, Nafisa Ghori, Stanley Falkow

Department of Microbiology and Immunology, School of Medicine, Stanford University, Stanford, California 94305, USA
e-mail: cdarby@stanford.edu

1. Bacot, A. W. & Martin, C. J. *J. Hygiene* **3** (suppl.), 423–439 (1914).

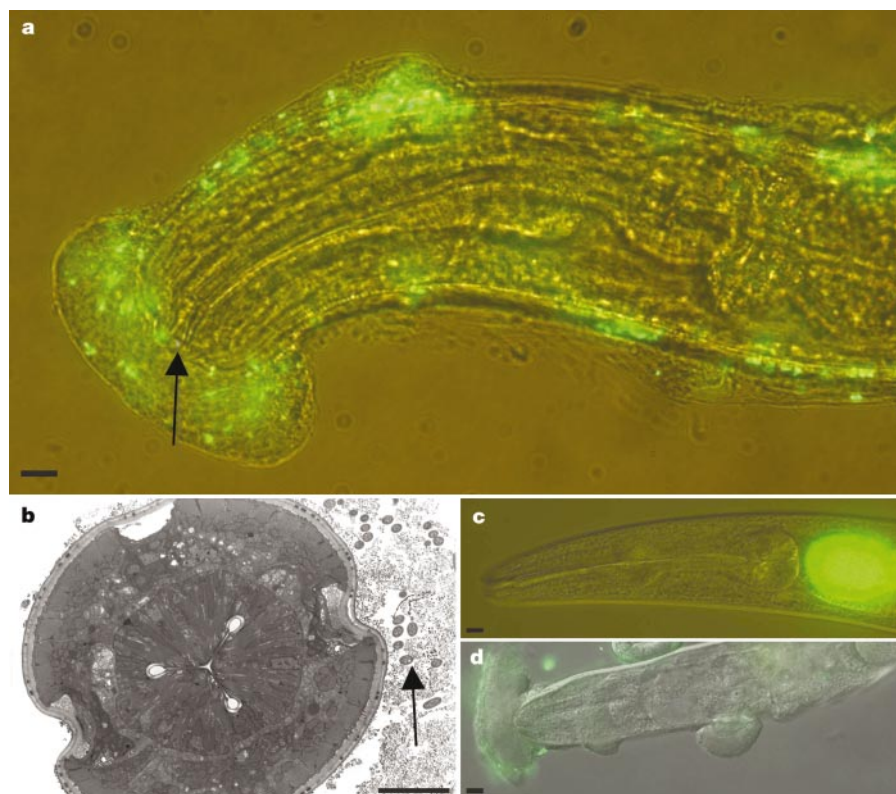


Figure 1 *Yersinia pseudotuberculosis* biofilms on *Caenorhabditis elegans*. **a**, *Y. pseudotuberculosis* expressing green fluorescent protein (GFP) embedded in a biofilm on the surface of *C. elegans*. Arrow shows the position of the worm's mouth. **b**, Transverse section through the *C. elegans* head. The biofilm has a granular appearance; arrow denotes a cluster of bacteria within the film. **c**, Control *C. elegans* without a biofilm after feeding on *Escherichia coli* expressing GFP; bright signal is in the anterior of the intestine. **d**, *C. elegans* after attempting to feed on *E. coli* expressing GFP in the presence of a biofilm produced by *Y. pseudotuberculosis*. Scale bars, 5 μ m.

2. Hinnebusch, B. J., Fischer, E. R. & Schwan, T. G. *J. Infect. Dis.* **178**, 1406–1415 (1998).
3. Hinnebusch, B. J., Perry, R. D. & Schwan, T. G. *Science* **273**, 367–370 (1996).
4. Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. *Annu. Rev. Microbiol.* **49**, 711–745 (1995).
5. Avery, L. *Genetics* **133**, 897–917 (1993).
6. Brubaker, R. R. *Clin. Microbiol. Rev.* **4**, 309–324 (1991).
7. Jones, H. A., Lillard, J. W. Jr & Perry, R. D. *Microbiology* **145**, 2117–2128 (1999).
8. Gerke, C., Kraft, A., Sussmuth, R., Schweitzer, O. & Gotz, F. *J. Biol. Chem.* **273**, 18586–18593 (1998).
9. Heilmann, C. et al. *Mol. Microbiol.* **20**, 1083–1091 (1996).
10. Hinnebusch, B. J. *J. Mol. Med.* **75**, 645–652 (1997).

Competing financial interests: none declared.

Phylogeny

A non-hyperthermophilic ancestor for Bacteria

The first phyla that emerge in the tree of life based on ribosomal RNA (rRNA) sequences are hyperthermophilic, which led to the hypothesis that the universal ancestor, and possibly the original living organism, was hyperthermophilic¹. Here we reanalyse the bacterial phylogeny based on rRNA using a more reliable approach, and find that hyperthermophilic bacteria (such as Aquificales and Thermotogales) do not emerge first, suggesting that the Bacteria had a non-hyperthermophilic ancestor. It seems that Planctomycetales, a phylum with numerous peculiarities, could be the first emerging bacterial group.

rRNA is a useful tool for investigating the universal phylogeny of life, particularly for uncultured organisms². The phylogeny of Bacteria based on rRNA is congruent

with genomic data, such as phylogenetic trees based on multiple genes^{3,4}. This suggests that rRNA is rarely transferred and indicates that species phylogeny can be deduced by using this marker.

However, rRNA phylogenies can be seriously affected by artefacts of tree construction⁵. As the most slowly evolving positions are less prone to confounding factors such as multiple substitutions, they probably retain an ancient phylogenetic signal⁶. Phylogenies constructed by examining these positions are less affected by artefacts. We used the 'slow-fast' method⁷ to identify these positions, which has proved efficient in analysing eukaryotic phylogeny based on rRNA⁵.

There are several discrepancies between the standard phylogeny² and the one we have deduced using the most conserved positions (Fig. 1). Our version shows hyperthermophilic Bacteria (Aquificales and Thermotogales) to be monophyletic, which is consistent with large-scale studies^{3,4}, instead of paraphyletic as in classical versions. More important, our phylogeny shows the late emergence of hyperthermophiles, as they are clustered with Fusobacteria within a vast multifurcation containing almost all the phyla.

Surprisingly, Planctomycetales emerge at the base of the Bacteria. Whereas the support for the early branching of hyperthermophilic Bacteria is low (bootstrap value about 20%), slowly evolving positions (Fig. 1) provide reasonable, albeit inconclusive, support for the early emergence of Planctomycetales (bootstrap value about 70%). The basal position of hyperthermophilic phyla is found only when examining noisy (that is,

fast-evolving) positions, and is thus probably artefactual. The artefactual attraction of the long branches⁸ of Archaea and hyperthermophilic Bacteria, and the high guanine and cytosine content of rRNA in most Archaea, Aquificales and Thermotogales⁹, could explain why hyperthermophiles are often found to branch early in Bacteria.

The emergence of hyperthermophilic Bacteria among mesophiles suggests a secondary adaptation to life at very high temperature for these organisms, which may have been facilitated by massive gene transfer from the Archaea¹⁰. The most convincing support for this idea is provided by the enzyme reverse gyrase, which is specific to hyperthermophiles¹¹. This enzyme, which introduces positive supercoils in DNA and is thought to stabilize it at high temperatures, has been independently acquired by *Aquifex* and *Thermotoga* from Archaea¹¹. Consistent with the ancestral guanine and cytosine content of rRNA⁹, our results indicate that the most recent common ancestor of Bacteria was not hyperthermophilic.

We find that Planctomycetales are the first branching bacterial group (Fig. 1). This is a significant, although understudied, division of Bacteria, whose members share several peculiar traits, such as a budding mode of reproduction and the lack of peptidoglycan in their cell walls¹². The most intriguing feature of this group is the existence of a single or double membrane around the chromosome in *Gemmata* sp. and *Pirellula* sp., respectively, which has been compared to the eukaryotic nucleus¹². However, evolutionary homology of these structures with the eukaryotic nucleus has not been proved. An early emergence of Planctomycetales, as inferred from our work, must be confirmed by a careful phylogenetic analysis of genome sequences from several apparently early-branching Bacteria. If our finding is verified, the origin of Bacteria should be seriously reconsidered.

Céline Brochier, Hervé Philippe

Phylogénie, Bioinformatique et Génome, UMR 7622 CNRS, Université Pierre et Marie Curie, 9 quai St Bernard, 75005 Paris, France e-mail : herve.philippe@snv.jussieu.fr

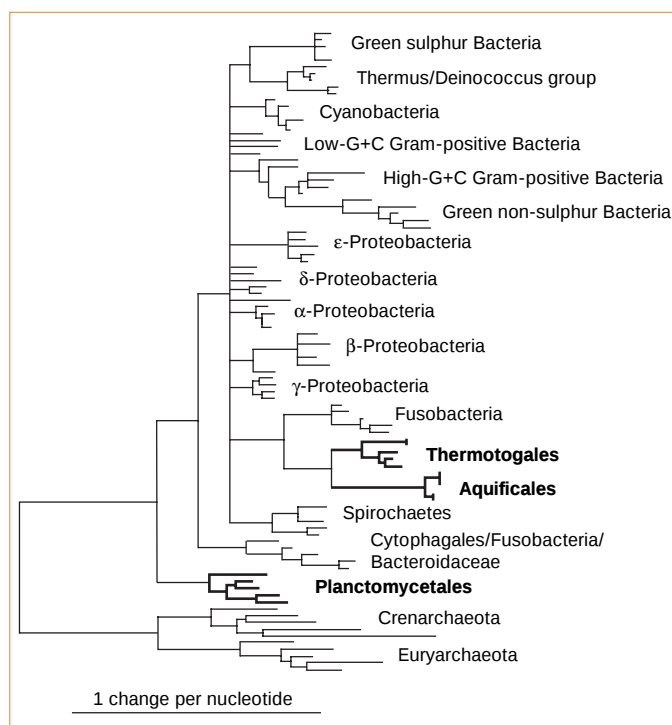


Figure 1 Prokaryotic phylogenetic tree based on conserved positions in ribosomal RNA. The 'slow-fast' method⁷ was used to estimate the rate of evolution at each position as the sum of the number of substitutions within 19 predefined phyla. Alignments were constructed for various thresholds (0–15) and the most parsimonious was inferred using PAUP 4b8 (alignments and trees are available from the authors). The relative positions of Planctomycetales, Aquificales and Thermotogales remain very similar for all thresholds from 1 to 10. The phylogeny shown here is based on the 751 positions with a threshold of fewer than 5 substitutions, which represents 45% of informative positions and is thus a valid compromise between the quality and the quantity of phylogenetic information.

1. Stetter, K. O. *Ciba Found. Symp.* **202**, 1–10 (1996).
2. Pace, N. R. *Science* **276**, 734–740 (1997).
3. Brochier, C., Baptiste, E., Moreira, D. & Philippe, H. *Trends Genet.* **18**, 1–5 (2002).
4. Daubin, V. & Gouy, M. *Genome Inform. Ser. Workshop Genome Inform.* **12**, 155–164 (2001).
5. Philippe, H., Germot, A. & Moreira, D. *Curr. Opin. Genet. Dev.* **10**, 596–601 (2000).
6. Felsenstein, J. *J. Mol. Evol.* **53**, 447–555 (2001).
7. Brinkmann, H. & Philippe, H. *Mol. Biol. Evol.* **16**, 817–825 (1999).
8. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
9. Galtier, N., Tourasse, N. & Gouy, M. *Science* **283**, 220–221 (1999).
10. Nelson, K. E. et al. *Nature* **399**, 323–329 (1999).
11. Forterre, P., Bouthier de la Tour, C., Philippe, H. & Duguet, M. *Trends Genet.* **16**, 152–154 (2000).
12. Fuerst, J. A. *Microbiology* **141**, 1493–1506 (1995).

Mechanism of glutamate receptor desensitization

Yu Sun^{*†}, Rich Olson^{*†}, Michelle Horning^{†‡}, Neali Armstrong^{*}, Mark Mayer[‡] & Eric Gouaux^{*§}

^{*} Department of Biochemistry and Molecular Biophysics and [§] Howard Hughes Medical Institute, Columbia University, 650 West 168th Street, New York, New York 10032, USA

[‡] Laboratory of Cellular and Molecular Neurophysiology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

[†] These authors contributed equally to this work

Ligand-gated ion channels transduce chemical signals into electrical impulses by opening a transmembrane pore in response to binding one or more neurotransmitter molecules. After activation, many ligand-gated ion channels enter a desensitized state in which the neurotransmitter remains bound but the ion channel is closed. Although receptor desensitization is crucial to the functioning of many ligand-gated ion channels *in vivo*, the molecular basis of this important process has until now defied analysis. Using the GluR2 AMPA-sensitive glutamate receptor, we show here that the ligand-binding cores form dimers and that stabilization of the intradimer interface by either mutations or allosteric modulators reduces desensitization. Perturbations that destabilize the interface enhance desensitization. Receptor activation involves conformational changes within each subunit that result in an increase in the separation of portions of the receptor that are linked to the ion channel. Our analysis defines the dimer interface in the resting and activated state, indicates how ligand binding is coupled to gating, and suggests modes of dimer–dimer interaction in the assembled tetramer. Desensitization occurs through rearrangement of the dimer interface, which disengages the agonist-induced conformational change in the ligand-binding core from the ion channel gate.

The canonical action of an agonist at a receptor involves activation of the receptor followed by either agonist unbinding—that is, deactivation—or receptor desensitization. Paradoxically, desensitization is characterized by entry into an inactive state, even though agonist remains tightly bound. Desensitization occurs in molecules as diverse as ion channels¹ and G-protein-coupled receptors² and participates in shaping the amplitude, duration and frequency of signalling within and between cells. Even though desensitization is ubiquitous in receptor-signal-transduction cascades, the molecular mechanisms underlying desensitization of ligand-gated ion channels are poorly understood.

First characterized by Katz and Thesleff during studies of the nicotinic acetylcholine receptor³, further experiments have shown that desensitization is a functionally important phenomenon¹ that occurs in most ligand-gated ion channels, and particularly in those at fast chemical synapses. Ionotropic glutamate receptors (iGluRs), crucial receptor molecules at chemical synapses, exhibit various rates and extents of desensitization depending on the receptor subtype⁴. Whereas *N*-methyl-D-aspartate (NMDA) receptors desensitize slowly via multiple mechanisms, α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) and kainate receptors desensitize profoundly on the millisecond timescale⁵. Over the past 20 years, a host of experiments performed on receptors for AMPA, kainate and NMDA have demonstrated that agonists, naturally occurring splice variants, engineered mutants, allosteric modulators and lectins modulate the rate and extent of desensitization⁴.

Studies on AMPA receptor desensitization with the use of hippocampal neurons and rapid-solution-exchange methods provided the first evidence for nearly complete desensitization in the presence of saturating concentrations of glutamate, AMPA and quisqualate^{6,7}. Subsequent experiments showed that desensitization is effectively promoted by glutamate concentrations much lower than required for activation, and recovery from desensitization proceeds at a rate at least 10-fold slower than that of deactivation^{8,9}. Accordingly, AMPA receptors desensitize from partly and fully liganded closed channel states, as well as from the open state¹⁰. A hallmark of the desensitized state is an increase in agonist affinity; in

AMPA receptors desensitization increases agonist affinity by about 20-fold¹¹. Even though numerous studies have probed desensitization from a functional perspective, a structural explanation of desensitization has proved elusive. Here we report structural and functional studies on the tetrameric^{12,13} AMPA-selective GluR2 receptor that localize crucial elements of the receptor involved in desensitization, illustrate how receptor activation is coupled to desensitization, allow us to develop quantitative relationships between the extent of receptor desensitization and the strength of intradimer subunit interfaces, and provide a molecular mechanism for the action of cyclothiazide (CTZ), a positive allosteric effector.

Blocking desensitization promotes dimerization

Similarly to native AMPA receptors in hippocampal neurons^{6,7}, homomeric GluR2 undergoes rapid and nearly complete desensitization when glutamate is applied; this is almost entirely blocked by the L483Y mutation¹⁴ or by the allosteric potentiator CTZ¹⁵, as indicated in Fig. 1. Both the L483Y site and alternatively spliced residues important for the action of CTZ, such as Ser 754, are contained within the sequence of the glutamate receptor ligand-binding core. As illustrated in Fig. 1b, this core is formed by segments S1 and S2 (ref. 16), which are separated by the transmembrane regions 1 and 2 and the pore loop (P). Covalent fusion of S1 and S2 yields the GluR2 S1S2J ligand-binding core, which reproduces the pharmacological behaviour of the full-length receptor and is a water-soluble construct suitable for biochemical and crystallographic studies^{17–19}.

We now show, using sedimentation equilibrium experiments, that the wild-type S1S2J (flop-N754) construct and the CTZ-sensitive S1S2J–N754S construct are essentially monomeric at protein concentrations below 1 mM. By contrast, the L483Y single-amino-acid mutant forms a dimer with a dissociation constant (K_d) of 30 nM, shifting the dimer-to-monomer K_d by 10^5 -fold relative to the wild-type S1S2J construct, as shown in Table 1. Similarly to the effect of the L483Y mutation, CTZ also promotes dimerization of S1S2J–N754S. Shown in Fig. 1c, d are plots of the centrifugation data for the S1S2J–L483Y and S1S2J–N754S proteins, respectively.

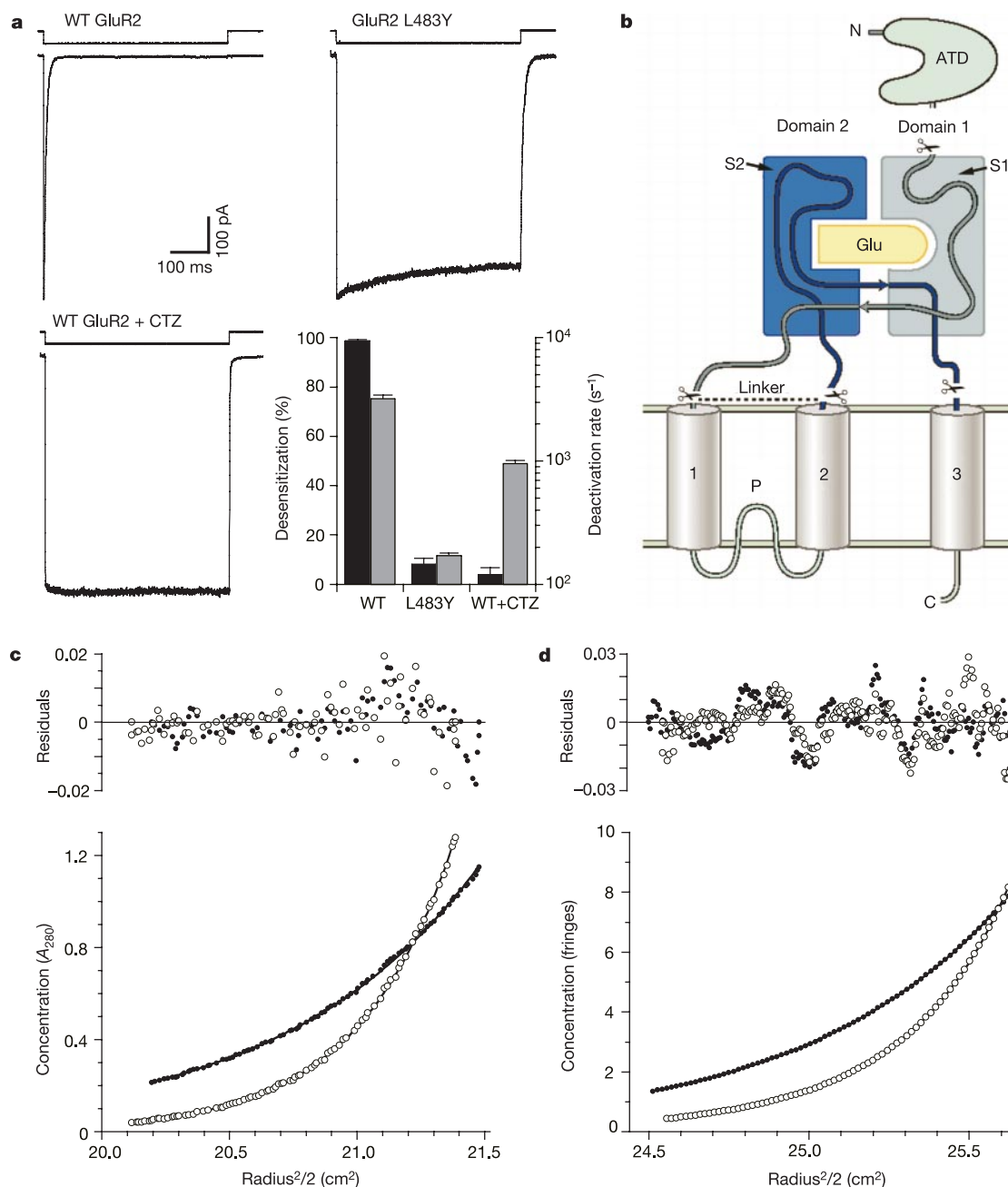


Figure 1 The L483Y mutation and CTZ promote dimerization and block desensitization. **a**, The rapid and nearly complete desensitization of the wild-type (WT) GluR2 receptor is blocked by either the L483Y mutation or by CTZ. In the bottom right panel, black columns show percentage desensitization and grey columns show deactivation rate. **b**, iGluR domain organization. Polypeptide segments S1 and S2 comprise the water-soluble ligand-binding core and the S1S2J construct studied here encompasses residues 392–506 (S1) and 632–775 (S2) linked together by a Gly–Thr dipeptide. The N-terminal domain (ATD) and the transmembrane segments are not contained within the S1S2J construct. **c**, The wild-type GluR2 S1S2J construct is predominately monomeric and the L483Y mutation markedly shifts the monomer–dimer equilibrium towards the dimer. Shown are sedimentation equilibrium scans of wild-type (filled circles) and L483Y (open circles) S1S2 constructs taken at 20,000 r.p.m. and a loading concentration of

0.5 mg ml⁻¹. Scans from three concentrations (0.25, 0.5 and 0.75 mg ml⁻¹) were globally fitted to a single-species model yielding a relative molecular mass (M_r) of 28,300 (28.3K) (95% confidence values of 27.0K and 29.6K) for wild-type and 58.2K (56.6K, 59.7K) for L483Y. The expected M_r of the monomer is 29.2K. **d**, CTZ also shifts the monomer–dimer equilibrium profoundly to the dimeric species. Sedimentation equilibrium scans of S1S2J–N754S in the presence and absence of 330 μM CTZ at 20,000 r.p.m. and a loading concentration of 1.5 mg ml⁻¹. Interference scans without water blanks are shown fitted to a single-species model for S1S2J–N754S alone (filled circles) and a monomer–dimer–tetramer equilibrium for S1S2J–N754S plus 330 μM CTZ (open circles). The M_r for S1S2J–N754S alone was calculated as 31.7K (31.3K, 32.0K) and the CTZ fit indicated a majority of the dimeric species with <10% monomer or tetramer.

L483Y dimer and CTZ complex structures

To define the mechanisms by which the L483Y mutant and CTZ promote dimerization, we performed a series of crystallographic experiments. Structural analysis of the L483Y mutant of the GluR2 S1S2J protein in complex with AMPA reveals a non-crystallographic

symmetry (NCS) related dimer, as illustrated in Fig. 2. The corresponding crystallographic statistics are shown in Table 2. Subunits in the dimer are arranged with their amino termini on the same surface facing ‘up’; the linker regions that replace the ion channel are on the opposite surface, proximal to the putative membrane plane.

Table 1 Dimer dissociation measurements and functional analysis of desensitization

Mutant	K_d (μ M)	ΔG_{dd} (kcal mol ⁻¹)	Desensitization (%)	ΔG_{des} (kcal mol ⁻¹)	k_{des} (s ⁻¹)	k_{rec} (s ⁻¹)	k_{off} (s ⁻¹)
L483Y	0.03	9.5	8.3 ± 2.2	1.38 ± 0.14	6.1 ± 0.8 (9)	n.d.	175 ± 7.7 (16)
L483F	0.68	7.8	38.9 ± 1.3	0.27 ± 0.03	18.1 ± 2.8 (6)	n.d.	374 ± 38 (8)
L483Y/K752A	0.89	7.6	26.1 ± 2.2	0.64 ± 0.07	17.1 ± 2.5 (15)	14.2 ± 3.5 (5)	590 ± 26 (17)
L483Y/L748A	0.52	7.9	42.8 ± 1.6	0.17 ± 0.04	13.3 ± 1.3 (13)	12.2 ± 4.1 (5)	100 ± 13 (13)
L483Y/S754D	5.33	6.7	52.0 ± 1.9	-0.05 ± 0.04	26.4 ± 2.6 (12)	10.8 ± n.d. (2)	268 ± 11 (11)
L483Y/L748A/K752A	7.22	6.5	90.1 ± 1.1	-1.33 ± 0.07	34.9 ± 2.3 (10)	20.1 ± 1.4 (7)	212 ± 14 (8)
Wild-type*	6,000	2.8	98.8 ± 0.3	-2.89 ± 0.16	185 ± 13 (13)	41.9 ± 4.8 (5)	3,107 ± 189 (12)
N754S (flip) + CTZ†	1.2	7.5	3.8 ± 0.8	2.16 ± 0.19	n.d.	n.d.	960 ± 54 (18)

K_d is the dimer dissociation constant and ΔG_{dd} is the ΔG of dimer dissociation calculated from the formula $\Delta G = -RT \ln K_d$ ($R = 1.987$ cal (mol K)⁻¹; $T = 277$ K). ΔG_{des} is the ΔG of desensitization calculated as described in Methods. k_{des} , k_{rec} and k_{off} are the rate of desensitization, rate of recovery from desensitization and rate of deactivation, respectively, measured by physiology and fast solution exchange as described in Methods. Measurements are the means ± s.e.m.; n.d. represents not determined. Numbers in parentheses are replicates.

* K_d for wild type is an estimate based on measurements of protein concentrations up to 8 mg ml⁻¹.

†This data set was fitted to an equilibrium model containing monomer, dimer and tetramer, resulting in a majority of the dimer species. By mass, the amount of each monomer and tetramer was less than 10% at a given concentration.

The dimer interface buries about 1,800 Å² of solvent-accessible surface area. Mediating all of the direct subunit–subunit contacts in the dimer are domain 1–domain 1 interactions between subunits A and B. Helix J, which is located near the carboxy terminus of the S2 segment and which includes one of the alternatively spliced flip/flop residues, as well as the Arg/Gly site of RNA editing²⁰, helix D, and the second interdomain β -strand, all participate in intersubunit contacts.

The L483Y dimer is knitted together by 2-fold related interactions between Tyr 483 on one subunit together with Leu 748 and Lys 752 on an adjacent subunit (Fig. 2e). The aromatic ring of Tyr 483 is sandwiched in a flat slot, making hydrophobic interactions with Leu 748 and the methylene groups of Lys 752. The proximity of the positively charged amino group of Lys 752 to the ring of Tyr 483 (~3.4 Å) permits favourable cation– π interactions²¹. Indeed, any aromatic group in the position equivalent to 483 in GluR2 greatly slows the rate and extent of desensitization in GluR3 (ref. 14), reinforcing the importance of side-chain shape and cation– π interactions.

Crystallographic analysis of the complex of GluR2 S1S2J–N754S with CTZ also reveals a 2-fold symmetric structure in which the organization and orientation of the subunits, as well as the dimer interface, are closely related to those seen in the L483Y structures (Fig. 2). Two CTZ molecules bind equivalently in the dimer interface with each molecule spanning the interface, displacing 10 ordered solvent molecules and burying ~480 Å² of solvent-accessible surface area each. The electron density reveals that the 1S-[1 α ,2 β (R*),4 α] isomer is preferentially bound to the receptor, which is consistent with functional studies identifying this isomer as being biologically the most active²². The hydrophobic bicyclic ring of molecule CTZ1 interacts with a complementary nonpolar pocket formed at the interface of the A and B subunits. The benzothiadiazide ring forms hydrogen bonds and nonpolar contacts with residues Pro 494 (A), Phe 495 (A) and Ser 497 (A) in β -strand 7, which links domains 1 and 2, as well as with Ser 754 (A) in helix J (ref. 18), as shown in Fig. 2f. Correspondingly, mutation of Ser 497 and Leu 751 weakens the action of CTZ²³.

Residue 754 is crucial in conferring CTZ sensitivity on AMPA receptors and in regulating the kinetics of desensitization²⁴. In flop

splice variants, residue 754 is asparagine and the receptors are relatively insensitive to CTZ. However, in flip splice forms of the receptor, residue 754 is serine and CTZ profoundly blocks desensitization²⁴. As seen in Fig. 2e, the asparagine side chain in the flop splice variant makes an intersubunit hydrogen bond to the backbone carbonyl oxygen of Ser 729 (refs 18, 19). When domain 1 of the S1S2J (flop)–Glu complex is superimposed on domain 1 of the S1S2J–N754S–Glu–CTZ structure, the distance between the asparagine side chain and CTZ is less than 2.5 Å, thus precluding the tight binding of CTZ. Destabilization of the dimer interface by the introduction of a negatively charged aspartic acid residue at position 754 has a marked effect on desensitization of the full-length receptor and subunit–subunit interactions in the S1S2J constructs, as discussed below.

Disruption of L483Y binding site

To show that the L483Y dimer interface has a key function in desensitization of the intact receptor, we performed the following experiments on wild-type and mutant receptors. We measured the K_d for dissociation of the water-soluble S1S2J dimers by equilibrium sedimentation in an analytical ultracentrifuge. In parallel, we assayed the electrical properties of the same mutants in the intact receptor by patch-clamp and rapid-solution-exchange methods and determined the extent, rate of onset (k_{des}) and recovery (k_{rec}) from desensitization, and the rate of receptor deactivation (k_{off}) as summarized in Table 1. Shown in Fig. 3 are representative traces of wild-type GluR2(flop) and the panel of site-directed mutants showing the response to rapid application of 3 mM L-glutamate for 1 and 500 ms. As seen in Fig. 3a, the wild-type receptor desensitizes nearly completely and deactivates very rapidly after the application of glutamate for 1 ms. By contrast, the L483Y mutant is refractory to desensitization¹⁴ and its rate of deactivation is 18-fold slower than that of the wild-type receptor.

As predicted from inspection of the L483Y and CTZ-complex crystal structures, removal of the Tyr 483 binding pocket or disruption of the CTZ-binding site destabilizes the S1S2J dimer in solution. When the residues interacting with Tyr 483 were removed successively, the resulting mutants formed weaker dimers than the L483Y parent (Table 1). Substituting a phenylalanine residue for

Table 2 Crystallographic refinement statistics

Protein	Ligand	Resolution (Å)	R_{work} (%)	R_{free} (%)	Average B value			Root-mean-square deviations			
					Overall	Main	Ligand	Bonds (Å)	Angles (deg)	B values	
										Bonds	Angles
L483Y	AMPA	20–2.3	21.5	26.5	22.80	21.85	12.54	0.006	1.18	2.1	3.0
L483Y	DNQX	20–2.3	22.0	26.8	16.26	15.25	21.80	0.006	1.19	2.0	2.8
N754D	Kainate	20–2.1	20.3	26.2	20.28	18.95	14.81	0.006	1.16	2.1	2.9
N754S	CTZ/Glu	30–1.8	22.2	24.3	20.63	19.23	17.64	0.006	1.33	1.8	2.7

$R_{work} = (\sum ||F_o| - |F_c||) / \sum |F_o|$, where F_o and F_c denote observed and calculated structure factors, respectively, and summation extends over all observed reflections; 10% of the data was set aside for calculation of R_{free} .

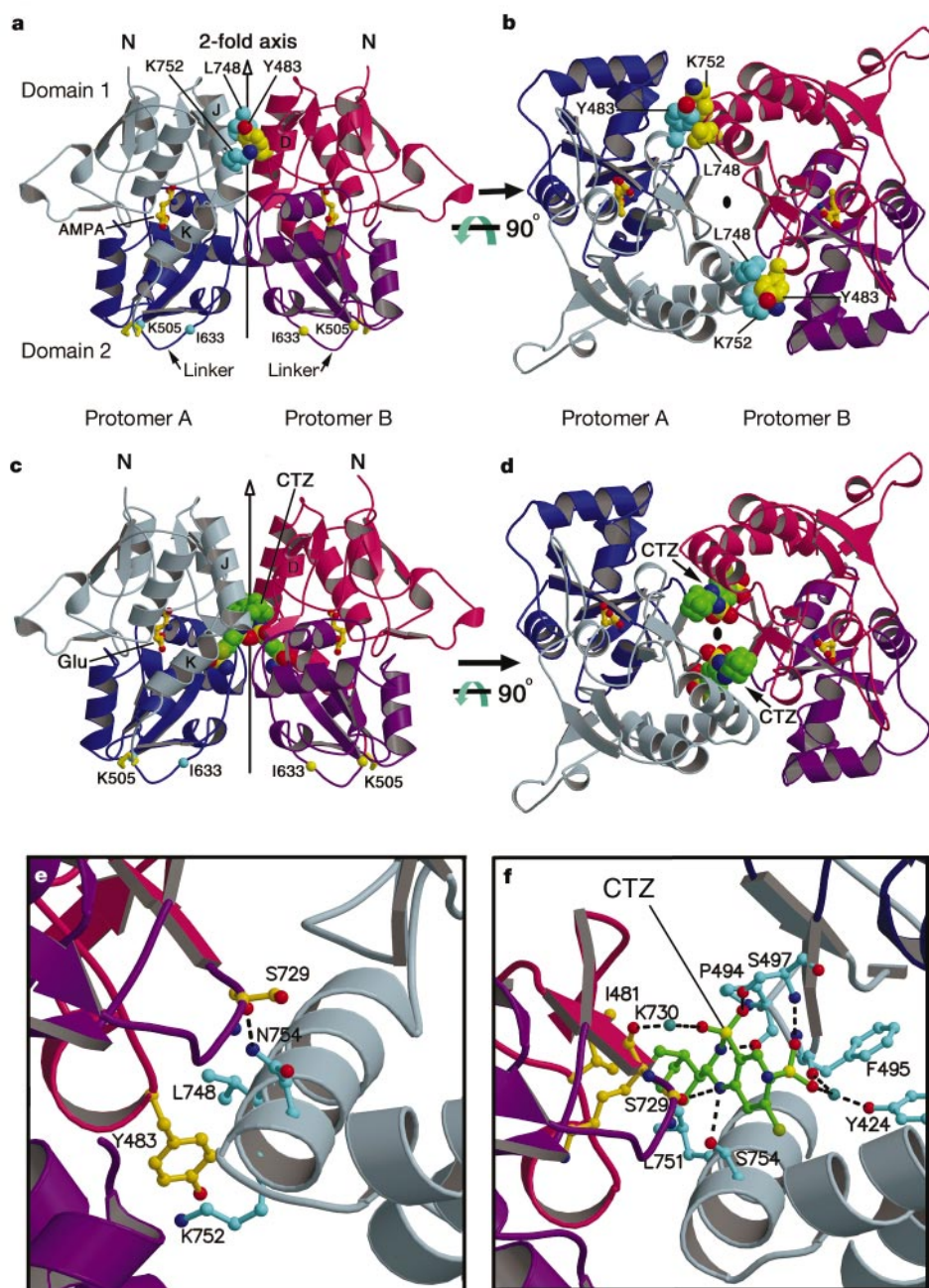


Figure 2 The L483Y mutation and CTZ stabilize the GluR2 S1S2J dimer. **a**, Side view of the S1S2J–L483Y dimer in complex with AMPA. Subunit A is grey (domain 1) and blue (domain 2). Subunit B is pink (domain 1) and purple (domain 2). Residues from A are cyan; residues from B are yellow. Lys 505 and Ile 633 flank transmembrane segments 1 and 2, respectively. **b**, Top view of the L483Y dimer looking down the 2-fold axis. **c**, CTZ stabilizes the GluR2 S1S2J–N754S dimer by binding in the dimer interface. Side view of the S1S2J dimer in a complex with glutamate and CTZ. The two CTZ molecules are green and are shown in CPK representation. **d**, Top view of the S1S2J–Glu–CTZ dimer, looking

down the 2-fold axis. **e**, Interactions between Tyr 483 from one subunit and Leu 748 and Lys 752 from another subunit. Similar interactions also occur in the dimer of S1S2J–L483Y in complex with DNQX. Note the intersubunit hydrogen bond between Asn 754 and the carbonyl oxygen of Ser 729. **f**, Interactions between CTZ and residues from subunits A (cyan) and B (yellow). The black dashed lines are hydrogen bonds and the light blue spheres are water molecules. Stereoviews of **e** and **f** are provided in Supplementary Information.

Leu 483 instead of tyrosine also destabilizes the dimer interface. Although we do not observe any direct protein–protein contacts between the hydroxyl group of Tyr 483 with the adjacent subunit, computational analysis suggests that phenylalanine forms weaker cation– π interactions than tyrosine²¹.

The decrease in S1S2J dimer stability is correlated with an increase in the desensitization of the intact receptor. The L483F mutant desensitizes substantially more than the L483Y mutant, although the rate of deactivation only doubles. Successive ablation

of the tyrosine acceptor site side chains increased desensitization without restoring the rate of deactivation to wild-type values, as illustrated in Fig. 3b. Our observation that the rate of deactivation slows 5–35-fold after the introduction of tyrosine or phenylalanine at position 483, independently of the effects of additional substitutions that produce large changes in the extent of desensitization, suggests that desensitization and deactivation are separate processes that might nevertheless share some common structural elements.

To quantify the extent of desensitization and relate it to the K_d of

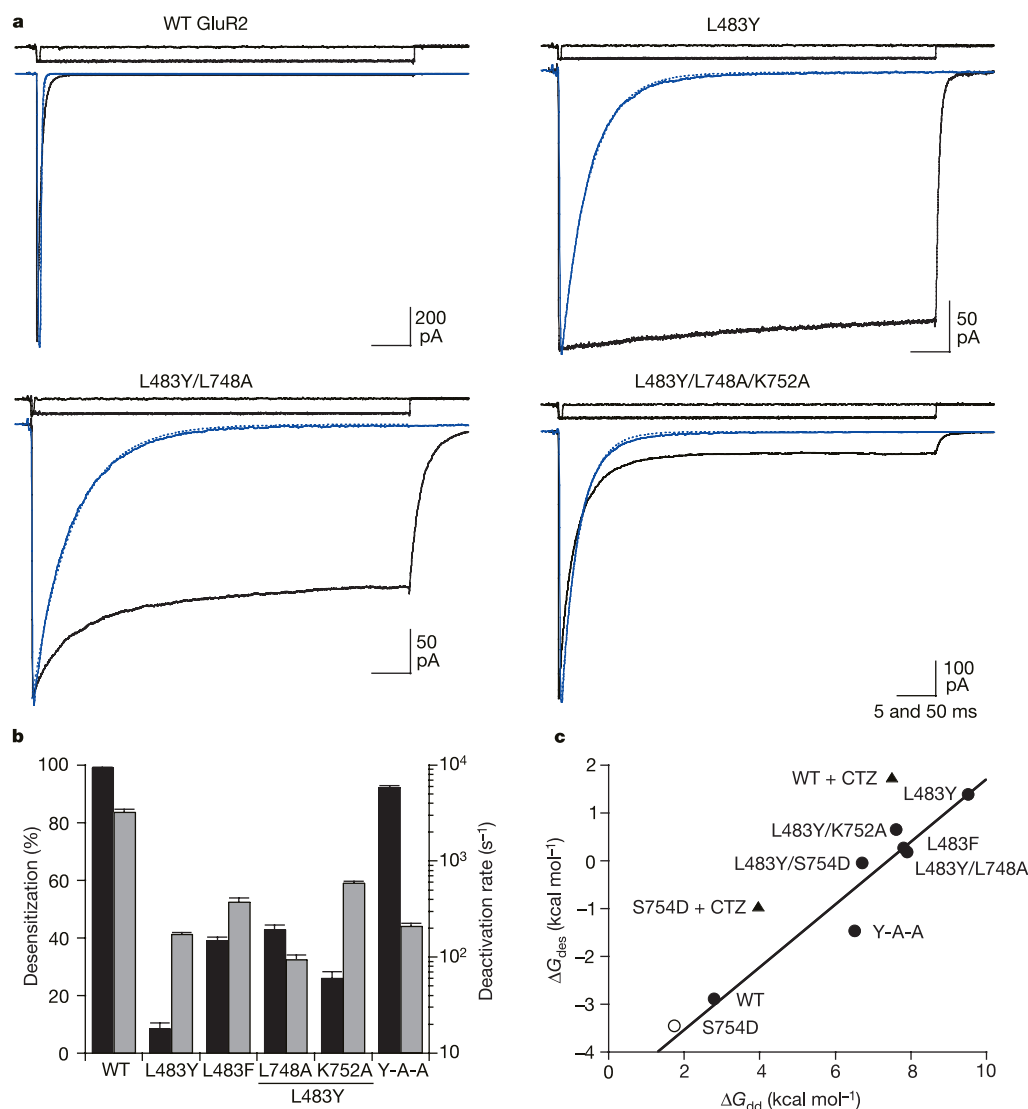


Figure 3 Disruption of the Tyr 483 binding site at the dimer interface increases the extent of receptor desensitization and shifts the S1S2J monomer–dimer equilibrium towards monomer. **a**, Superimposed responses to 1-ms and 500-ms applications (blue and black traces respectively) of 3 mM L-glutamate recorded from outside-out patches of HEK 293 cells transfected with the indicated GluR2 subunit cDNAs. For each panel the upper traces show open tip responses recorded at the end of the experiment; lines through the data points are single-exponential fits; the time calibrations for responses to 1-ms and 500-ms applications of glutamate differ 10-fold. **b**, Restoration of desensitization (black bars) for L483Y binding site mutants occurs without a return to the fast deactivation rate characteristic of wild-type (WT) GluR2 (grey bars); plots show means $\pm$ s.e.m. for observations recorded from 8–16 patches per cDNA. **c**, Correlation of the free-energy change for desensitization, ΔG_{des} , with that for dimerization of ligand-binding core mutants, ΔG_{dd} . Y-A-A indicates the triple mutant L483Y/L748A/K752A; the line was fitted by linear regression ($r = 0.95$). The points for the wild-type receptor and the S754D mutant with CTZ are displaced equally to the left and were not included in the fit. The open circle indicates the predicted ΔG_{dd} for the S1S2J–S754D mutant, extrapolated from the measured ΔG_{des} .

dimer dissociation for the GluR2 S1S2J constructs, we defined the ΔG of desensitization as

$$\Delta G_{des} = -RT \ln[A_2D_2]/[A_2O_2]$$

where $[A_2D_2]$ and $[A_2O_2]$ are the concentrations of the desensitized and activated dimers in full-length receptors at steady state, respectively, and determined as described in Methods. The ΔG of dimer dissociation (ΔG_{dd}) for the GluR2 S1S2J constructs was calculated from the respective K_d values. As illustrated in Fig. 3c, this analysis reveals a remarkably linear correlation between the extent of desensitization (ΔG_{des}) and the stability of the dimer interface (ΔG_{dd}).

Perturbation of dimer interface

According to the model described above, destabilization of the

dimer interface should enhance desensitization of the receptor, increase the dimer K_d and disrupt the dimer interface seen in the crystal. To test this hypothesis we took advantage of the fact that mutation of residue 754 to aspartic acid in GluR1 produced receptors that are active only in the presence of CTZ, indicating either high occupancy of the desensitized state in the absence of agonist²⁵ or an acceleration of the rate of desensitization such that after the binding of agonist a high proportion of receptors desensitize before the channel can open. As shown in Fig. 4, the GluR2 S754D mutant is indeed activated by glutamate, but desensitizes 12-fold more rapidly than the wild-type receptor. GluR2 S754D also shows weaker modulation by either the L483Y mutation or the addition of CTZ. However, similarly to the wild-type receptor, the rate of desensitization of the S754D mutant is slowed 95-fold after introduction of the L483Y mutation and 180-fold after the addition

of CTZ, suggesting that the S754D and L483Y mutations act independently, which is consistent with their being located at different sites on the dimer interface. Because the rate of recovery from desensitization for the S754D mutant is only 2.8-fold slower than that for wild-type GluR2, we suggest that there is only a moderate increase in the stability of the desensitized state relative to the wild-type receptor. Illustrated in Fig. 4b are the extents and rates of desensitization of the wild-type and S754D mutants. Accompanying the increase in rate and extent of desensitization produced by the S754D mutation is an increase in the dimer K_d for the S1S2J–L483Y/N754D ligand-binding core, relative to the S1S2J–L483Y species (Table 1).

Dimer–dimer interactions

Crystallographic analysis of the GluR2 S1S2J–N754D mutant showed no evidence of the dimer interface seen before in the wild-type and L483Y structures, which is almost certainly due to the replacement of an intersubunit hydrogen bond by a repulsive interaction between the N754D mutant side chain and the peptide-bond carbonyl oxygen of S729. Instead, we find a new subunit–subunit interaction in which one subunit is related to the second by a translation. This lateral interface involves loop 1 and helix K (domain 1) of protomer A and helices F and G (domain 2) of protomer C, as illustrated in Fig. 4c. The polar subunit A–subunit C interface buries 511 Å² of solvent-accessible surface area, consists of

a series of direct protein–protein contacts and water-mediated hydrogen bonds and includes a number of residues that are conserved in all non-NMDA receptors. The lateral interface formed between subunits A and C indicates a mechanism by which dimeric units^{19,26} within a glutamate receptor tetramer might interact (see Supplementary Information). Supporting the contention that loop 1 in GluR2 is involved in subunit–subunit interactions are results from work on the NMDA receptor indicating that a region equivalent to loop 1 participates in intersubunit contacts between NR1 and NR2 subunits²⁷. In the intact GluR2 receptor, subunits occupying the positions of A and C might be related by a translation and a rotation, in part because there is no evidence suggesting that glutamate receptors form extended ribbon-like aggregates.

Symmetry mismatch

Numerous independent lines of investigation indicate that the channel-forming region of iGluRs forms a roughly 4-fold rotationally symmetric tetramer similar to that of potassium channels^{28–32}. Thus, there must be a symmetry mismatch³³ between the ligand-binding cores, which have local rotational 2-fold axes of symmetry, and the transmembrane ion channel, which has an approximate local 4-fold axis of rotational symmetry. A possible mechanism to account for the symmetry transformation between the ligand-binding cores and the ion channel is described below and in Supplementary Information. Viewed from the ligand-binding

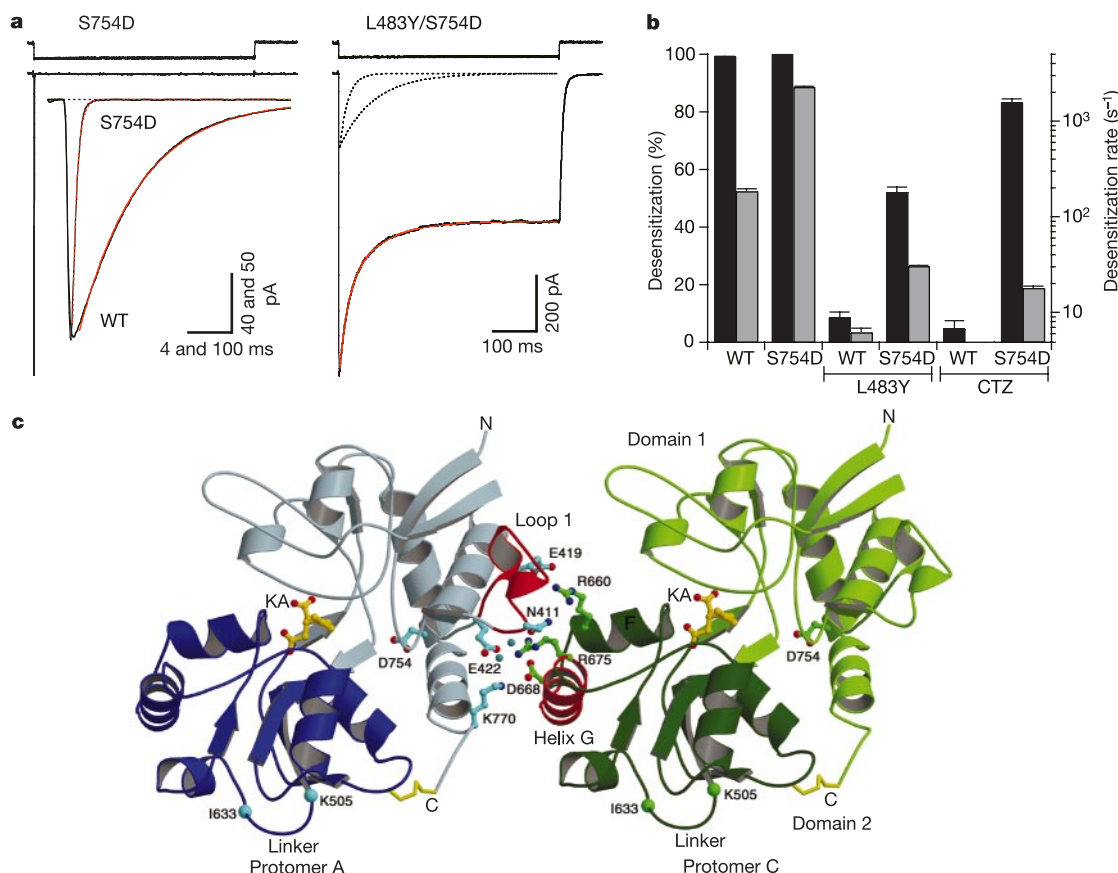


Figure 4 Introduction of aspartate at position 754 accelerates desensitization and reveals a new, 'lateral' mode by which the ligand-binding cores can interact. **a**, Responses to 500-ms applications of 3 mM L-glutamate recorded from outside-out patches of HEK 293 cells transfected with the indicated GluR2 subunit cDNAs; the onset of desensitization is fitted with one exponential function (left panel) or the sum of two exponential functions (right panel) (red lines), the individual terms of which are shown as dotted lines. For the left panel the superimposed responses on a faster time for wild-type (WT) GluR2 and S754D are inset. **b**, Analysis of the extent (black bars) and rate of onset of desensitization (grey

bars) for responses recorded from 13–25 patches per condition reveals that the S754D mutant attenuates the block of desensitization by both the L483Y mutant and CTZ; error bars show means $\pm$ s.e.m. **c**, The Asp 754 S1S2J fragment adopts a new packing in the crystal. Shown are subunits A (grey, domain 1; blue, domain 2) and C (light green, domain 1; dark green, domain 2), which are related by a translation. The interface is formed by loop 1 (red) and helix K of subunit A and helix G (red) of subunit C. Residues interacting across the interface are cyan (A) and green (C). Blue spheres are water molecules.

cores, towards the membrane, the polypeptide segments in subunit A that bridge the ligand-binding core and the transmembrane channel undergo a relative clockwise rotation of 45° and the segments bridging the same region in subunit B undergo an anti-clockwise rotation of 45° . The symmetry mismatch and local asymmetry have at least two important consequences. First, the extent to which ligand binding is coupled to the transmembrane channel might also be different for each subunit in a dimeric unit. Second, if we assume that there is a global 2-fold axis, then

diagonally related subunits occupy equivalent positions in the extracellular portion of the receptor and in the transmembrane channel region. Although symmetry mismatch and molecular asymmetry are relatively uncommon, there are several proteins, including picornavirus capsids, pyruvate dehydrogenase and HIV (human immunodeficiency virus) protease, that display such properties³³.

Activation

The evidence from the structural, biophysical and physiological experiments described above supports the hypothesis that the ligand-binding cores are organized as dimeric units in a tetrameric assembly^{19,34,35}. To understand how domain closure induced by agonist binding to an individual ligand-binding core alters the conformation of the domains in S1S2J dimers, we solved and compared the structures of the non-desensitizing S1S2J–L483Y mutant crystallized in the presence of either the agonist AMPA or the competitive antagonist DNQX (Fig. 5). Previous studies on wild-type GluR2 S1S2J have shown that DNQX stabilizes the ligand-binding core in an open, apo-like conformation and that AMPA and glutamate stabilize a closed-cleft conformation¹⁹. In the L483Y and wild-type S1S2J complexes, the dimer interface does not rearrange in response to domain closure, as shown in Fig. 5, because the dimer interface is formed exclusively by contacts in domain 1 and there are no intersubunit contacts involving domain 2, which is thus free to close in the agonist-bound complex. Indeed, superposition of the wild-type S1S2J–AMPA dimer¹⁹ and the nondesensitizing L483Y–AMPA dimer using α -carbon atoms from both subunits gives a root-mean-square deviation value of 0.47 Å, indicating that the dimer interface in both structures is representative of the active rather than the desensitized state.

To determine the consequence of agonist-induced domain closure on the L483Y dimer, we focus on Ile 633, which is adjacent to the linker between S1 and S2 and proximal to transmembrane segment 2. On binding a full agonist such as AMPA, the α -carbon atoms of Ile 633 on subunits A and B separate by 8 Å, undergo a vertical translation of 2.5 Å away from the putative membrane plane and rotate 1.25° around the 2-fold axis. We suggest that the increase in the separation of the linker regions is directly coupled to activation of the transmembrane ion channel and that this ‘coupling’ requires a stable dimer interface. If the dimer interface is prevented from rearrangement by the L483Y mutation or CTZ, then domain closure can be translated to an increase in the separation of residues proximal to Ile 633 and, in turn, opening of the ion channel gate(s). However, a reorientation of the dimer interface would decouple domain closure from linker separation and thus from ion channel gating. In this situation, which is what we suggest happens in the wild-type receptor on desensitization, agonist binds to the receptor and triggers domain closure, but no conformational changes are transmitted to the ion channel gate because the dimer interface has rearranged.

Mechanistic scheme

Illustrated in Fig. 5 is a mechanism for the activation and desensitization of the GluR2 AMPA-sensitive receptor in which we focus on the functional states and conformational changes that occur in a dimeric half of the tetrameric receptor. We concentrate on three regions of the receptor: the dimer interface, the domain 1–domain 2 ‘clamshell’ ligand-binding cleft and the gate of the ion channel. In the resting, apo, state of the receptor (A_0C_2), the dimer interface adopts a configuration like that seen in the L483Y and CTZ-complex structures. When agonists bind to the receptor, the initial interaction probably occurs via domain 1 (A_2C_2)^{19,36}. Subsequently, the ligand-binding domain closes and traps the agonist in the cleft between domains 1 and 2. On activation, the dimer interface remains intact and the conformational strain caused by domain closure is transferred to the gate, thus opening the channel (A_2O_2).

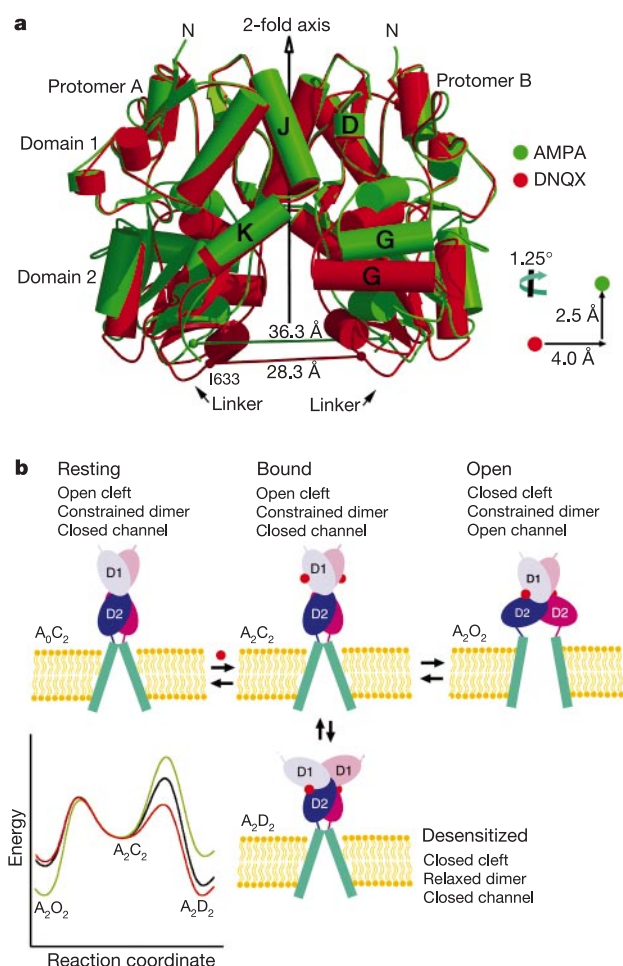


Figure 5 Agonist-induced conformational changes in the dimer and gating model. **a**, Overlap of the S1S2J–L483Y dimers bound with either an agonist (AMPA, green) or an antagonist (DNQX, red). The relative movement of the linker region, which connects the ligand-binding core to the channel-forming segments, is represented by the difference in position of Ile 633 in the two structures. Distances between Ile 633 on two protomers are 28.3 Å in the DNQX structure and 36.3 Å in the AMPA structure. In addition, Ile 633 rotates around the 2-fold axis by 1.25° and moves 2.5 Å along the 2-fold axis, away from the membrane. **b**, A model for glutamate receptor activation and desensitization. Domain 1 and domain 2 of the ligand-binding core are labelled D1 and D2, respectively. Transmembrane segments of each subunit are indicated by a single green cylinder and the N-terminal domain (ATD) has not been included in the model. Each subunit binds a single agonist (A, red circle) and exists in three distinct conformations: closed (C), open (O) and desensitized (D). The closed and open states share the same S1S2 dimer interface. After the binding of agonist, closure of domain 2 towards domain 1 opens the channel gate, whereas closure of domain 1 towards domain 2 disrupts the dimer interface and desensitizes the receptor. The states are connected by using a simplified model for activation and desensitization, more complex versions of which quantitatively describe AMPA receptor responses^{10,25}. A hypothetical plot of the free-energy change occurring during activation and desensitization is shown in the lower left panel for the wild-type (black line), L483Y (green line) and S754D (red line) species.

On desensitization, the dimer interface rearranges and therefore domain closure is decoupled from the channel gate (A_2D_2). For the wild-type receptor, the energy barrier for desensitization is higher than for activation and therefore the receptor activates faster than it desensitizes. However, the desensitized receptor is more stable than the activated receptor, and thus during a prolonged incubation with agonist most of the receptors become desensitized. Because desensitization is controlled by the interface between subunits, it is likely that desensitization involves a concerted conformational change within a dimeric unit.

Here we present structural, biophysical and electrical measurements to support a mechanism for the activation and desensitization of glutamate receptors. This mechanism explains how the work done by agonist-induced closure of the ligand-binding cleft is coupled to activation of the ion channel and to desensitization of the receptor. Comparison of the properties of the non-desensitizing, partly desensitizing and fully desensitizing mutants of GluR2 indicates that the coupling of agonist binding and channel gating is controlled by the stability of the dimer interface. Our studies provide strong support for the organization of tetrameric glutamate receptors as a pair of dimers in which there is a symmetry mismatch between the ligand-binding cores and the transmembrane channel. The lateral interface observed in the crystal structure of N754D provides new insight into how the dimers might be organized in the receptor and indicates a role for loop 1 on the interactions between dimers. The binding pocket of CTZ in the dimer interface explains for the first time a mechanism by which iGluRs can be modulated, which might also be useful for understanding the function of other allosteric modulators of iGluRs³⁷. □

Methods

Structure determination

AMPA and DNQX–S1S2J–L483Y and kainate–S1S2J–N754D data sets were collected with Cu K α radiation and an R-Axis IV area detector. The N754S–Glu–CTZ data set was collected at the National Synchrotron Light Source beam line x4A using a Quantum 4 charged-coupled device detector. All data sets were indexed, scaled and merged with HKL³⁸. The AMPA and DNQX–S1S2J–L483Y structures were determined by molecular replacement with the use of the structures of S1S2J complexed with the corresponding ligand as the search probe¹⁹, and the program AmoRe³⁹. The kainate–S1S2J–N754D structure was determined by molecular replacement with the use of the S1S2J–kainate structure¹⁸. After an initial round of rigid-body refinement, all three structures were subjected to a slow-cool simulated-annealing run starting at 4,000 K to reduce model bias. Iterative rounds of model building into omit maps and individual B -value refinements were performed until the R_{free} value converged. Ligands were modelled into the $F_o - F_c$ maps when R_{free} decreased below 0.30. Because the N754S–Glu–CTZ crystals were isomorphous with the S1S2J glutamate crystals, structure determination began with rigid-body refinement of the S1S2J–Glu structure with all the water molecules removed from the dimer interface. After one cycle of rigid-body refinement, Powell minimization and individual B -value refinement in the program CNS⁴⁰, R_{work} was 0.245 and R_{free} was 0.285. $F_o - F_c$ maps revealed clear density for two CTZ molecules in the dimer interface. The initial CTZ model was generated and energy-minimized by ChemDraw3D. The stereochemistry of CTZ was determined from the electron density maps and the model was modified accordingly. The program O was used for model building and map display⁴¹; LSQMAN was used to superimpose structures⁴². All other crystallographic calculations were done in the CCP4 program suite⁴³. For data collection statistics see Supplementary Information.

Sedimentation equilibrium

Sedimentation equilibrium runs were performed in a Beckman/Coulter analytical ultracentrifuge. S1S2J mutant analysis used the absorbance optical system at 229 or 280 nm with quartz or sapphire windows, respectively. Purified protein was dialysed extensively against a buffer containing 20 mM sodium phosphate pH 7.5, 150 mM NaCl, 1 mM EDTA and 1 mM L-glutamate and loaded into six-sector equilibrium cells at 0.75, 0.5 and 0.25 mg ml⁻¹ or 0.05, 0.02 and 0.005 mg ml⁻¹ depending on the dissociation constant. Samples were centrifuged at 15,000, 20,000 and 27,000 r.p.m. at 4 °C and scans were taken at 1-h intervals with a 0.001-cm spacing and ten replicates per point. Owing to the absorbance of CTZ at 280 nm, the Rayleigh interference system was used with an exterior-loading cell equipped with sapphire windows⁴⁴. These samples were run at ~1.5 mg ml⁻¹ with water blanks unless otherwise noted. Data analysis was performed with the WinNONLIN least-squares fitting program⁴⁵ as described previously²⁶. For additional statistics for the centrifugation experiments see Supplementary Information.

Physiology

HEK 293 cells grown on 35-mm dishes were transfected with 3 μ g complementary DNA⁴⁶.

Recordings were performed at room temperature with an Axopatch 200A amplifier and outside-out patches positioned in front of four-bore glass tubing mounted on a P245.30 piezoelectric stack driven by a P-270 HVA amplifier (Physik Instrumente). The external solution contained (in mM): 145 NaCl, 2.5 KCl, 1.8 CaCl₂, 1 MgCl₂, 5 HEPES pH 7.3 and 10 glucose; 7 mM NaCl was added to the barrel containing 3 mM L-glutamate to increase the amplitude of open-tip responses. The internal solution contained (in mM): 110 CsF, 35 CsCl, 5 Cs₄BAPTA, 0.5 CaCl₂, 1.0 MgCl₂, 10 HEPES pH 7.2. Data were recorded with a 16-bit analogue-to-digital interface under control of Synapse software on a Macintosh G4 computer; analysis was performed with Synapse and Kaleidagraph.

To calculate the ΔG of desensitization (ΔG_{des}) we assumed that at the peak of inward current responses all ligand-binding sites are occupied by glutamate and that, at equilibrium, dimers in the open state (A_2O_2) are in equilibrium with dimers in the desensitized state (A_2D_2). Here we make the simplifying assumption that the minimal functional unit within the tetrameric receptor is a dimer. ΔG_{des} is defined as follows

$$\Delta G_{\text{des}} = -RT \ln[A_2D_2]/[A_2O_2]$$

By definition, percentage desensitization equals $100 \times (1 - I_{\text{ss}}/I_{\text{peak}})$, where I_{ss} and I_{peak} are the amplitudes of currents at steady state and peak, respectively, and $[R_o]_{\text{ss}}$ and $[R_o]_{\text{peak}}$ are the corresponding populations of receptors in the open state. Assuming that all receptors open at the peak response, that is, $[R_o]_{\text{peak}} = [R_{\text{total}}]$, then $I_{\text{ss}}/I_{\text{peak}} = [R_o]_{\text{ss}}/[R_{\text{total}}]$. Thus

$$\text{percentage desensitization} = 100 \times (1 - [R_o]_{\text{ss}}/[R_{\text{total}}]).$$

At equilibrium, the total number of receptors ($[R_{\text{total}}]$) equals the sum of the population of the desensitized $[R_d]_{\text{ss}}$ and open ($[R_o]_{\text{ss}}$) receptor populations.

Therefore

$$[A_2D_2]/[A_2O_2] = [R_d]_{\text{ss}}/[R_o]_{\text{ss}}$$

$$= ([R_{\text{total}}] - [R_o]_{\text{ss}})/[R_o]_{\text{ss}}$$

$$= 1/(1 - \text{percentage desensitization}/100) - 1$$

Received 24 January; accepted 4 April 2002.

- Jones, M. V. & Westbrook, G. L. The impact of receptor desensitization on fast synaptic transmission. *Trends Neurosci.* **19**, 96–101 (1996).
- Clark, R. B., Knoll, B. J. & Barber, R. Partial agonists and G-protein coupled receptor desensitization. *Trends Pharmacol. Sci.* **20**, 279–286 (1999).
- Katz, B. & Thesleff, S. A study of the 'desensitization' produced by acetylcholine at the motor end-plate. *J. Physiol. (Lond.)* **138**, 63–80 (1957).
- Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* **51**, 7–61 (1999).
- Mayer, M. L. & Partin, K. M. in *Excitatory Amino Acids and Synaptic Function* (eds Wheal, H. & Thomson, A.) 89–98 (Academic, New York, 1995).
- Kiskin, N. I., Krishtal, O. A. & Tsyndrenko, A. Y. Excitatory amino acid receptors in hippocampal neurons: kainate fails to desensitize them. *Neurosci. Lett.* **63**, 225–230 (1986).
- Mayer, M. L. & Vyklicky, L. Concanavalin A selectively reduces desensitization of mammalian neuronal quisqualate receptors. *Proc. Natl Acad. Sci. USA* **86**, 1411–1415 (1989).
- Trussell, L. O., Thio, L. L., Zorumski, C. F. & Fischbach, G. D. Rapid desensitization of glutamate receptors in vertebrate central neurons. *Proc. Natl Acad. Sci. USA* **85**, 4562–4566 (1988).
- Colquhoun, D., Jonas, P. & Sakmann, B. Action of brief pulses of glutamate on AMPA/kainate receptors in patches from different neurons of rat hippocampal slices. *J. Physiol. (Lond.)* **458**, 261–287 (1992).
- Hausser, M. & Roth, A. Dendritic and somatic glutamate receptor channels in rat cerebellar Purkinje cells. *J. Physiol. (Lond.)* **501**, 77–95 (1997).
- Hall, R. A., Kessler, M., Quan, A., Ambros-Ingerson, J. & Lynch, G. Cyclothiazide decreases [³H]AMPA binding to rat brain membranes: evidence that AMPA receptor desensitization increases agonist affinity. *Brain Res.* **628**, 345–348 (1993).
- Rosenmund, C., Stern-Bach, Y. & Stevens, C. F. The tetrameric structure of a glutamate receptor channel. *Science* **280**, 1596–1599 (1998).
- Laube, B., Kuhse, J. & Betz, H. Evidence for a tetrameric structure of recombinant NMDA receptors. *J. Neurosci.* **18**, 2954–2961 (1998).
- Stern-Bach, Y., Russo, S., Neuman, M. & Rosenmund, C. A point mutation in the glutamate binding site blocks desensitization of AMPA receptors. *Neuron* **21**, 907–918 (1998).
- Patneau, D. K., Vyklicky, L. & Mayer, M. L. Hippocampal neurons exhibit cyclothiazide-sensitive rapidly desensitizing responses to kainate. *J. Neurosci.* **13**, 3496–3509 (1993).
- Stern-Bach, Y. *et al.* Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* **13**, 1345–1357 (1994).
- Kuusinen, A., Arvola, M. & Keinänen, K. Molecular dissection of the agonist binding site of an AMPA receptor. *EMBO J.* **14**, 6327–6332 (1995).
- Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
- Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: Crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).
- Sommer, B. *et al.* Flip and flop: A cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580–1585 (1990).
- Gallivan, J. P. & Dougherty, D. A. Cation- π interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
- Cordi, A. A. *et al.* Identification and characterization of the isomers of cyclothiazide responsible for potentiating AMPA current. *Bioorg. Med. Chem. Lett.* **4**, 1957–1960 (1994).
- Partin, K. M. Domain interactions regulating AMPA receptor desensitization. *J. Neurosci.* **21**, 1939–1948 (2001).
- Partin, K. M., Bowie, D. & Mayer, M. L. Structural determinants of allosteric regulation in alternatively spliced AMPA receptors. *Neuron* **14**, 833–843 (1995).
- Partin, K. M., Fleck, M. W. & Mayer, M. L. AMPA receptor flip/flop mutants affecting deactivation,

- desensitization, and modulation by cyclothiazide, aniracetam, and thiocyanate. *J. Neurosci.* **16**, 6634–6647 (1996).
26. Mayer, M. L., Olson, R. & Gouaux, E. Mechanisms for ligand binding to GluR0 ion channels: crystal structures of the glutamate and serine complexes and a closed apo state. *J. Mol. Biol.* **311**, 815–836 (2001).
27. Regalado, M. P., Villarreal, A. & Lerma, J. Intersubunit cooperativity in the NMDA receptor. *Neuron* **32**, 1085–1096 (2001).
28. Wo, Z. G. & Oswald, R. E. Unraveling the modular design of glutamate-gated ion channels. *Trends Neurosci.* **18**, 161–168 (1995).
29. Wood, M. W., VanDongen, H. M. & VanDongen, A. M. Structural conservation of ion conduction pathways in K channels and glutamate receptors. *Proc. Natl Acad. Sci. USA* **92**, 4882–4886 (1995).
30. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
31. Chen, G.-Q., Cui, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).
32. Panchenko, V. A., Glasser, C. R. & Mayer, M. L. Structural similarities between glutamate receptors channels and K⁺ channels examined by scanning mutagenesis. *J. Gen. Physiol.* **117**, 345–360 (2001).
33. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
34. Kuusinen, A., Abele, R., Madden, D. R. & Keinänen, K. Oligomerization and ligand-binding properties of the ectodomain of the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor subunit GluRD. *J. Biol. Chem.* **274**, 28937–28943 (1999).
35. Ayolan, G. & Stern-Bach, Y. Functional assembly of AMPA and kainate receptors is mediated by several discrete protein–protein interactions. *Neuron* **31**, 103–113 (2001).
36. Abele, R., Keinänen, K. & Madden, D. R. Agonist-induced isomerization in a glutamate receptor ligand-binding domain: a kinetic and mutagenic analysis. *J. Biol. Chem.* **275**, 21355–21363 (2000).
37. Nagarajan, N., Quast, C., Boxall, A. R., Shahid, M. & Rosenmund, C. Mechanism and impact of allosteric AMPA receptor modulation by the ampakine CX546. *Neuropharmacology* **41**, 650–663 (2001).
38. Otwinowsky, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
39. Navaza, J. *AMoRe*: An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
40. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
42. Kleywegt, G. J. & Jones, T. A. A super position. *CCP4/EST-EACBM Newsl. Prot. Crystallogr.*, 9–14 (1994).
43. Collaborative Computational Project, Number 4 The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Ansevin, A. T., Roark, D. E. & Yphantis, D. A. Improved ultracentrifuge cells for high-speed sedimentation equilibrium studies with interference optics. *Anal. Biochem.* **34**, 237–261 (1970).
45. Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588 (1981).
46. Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7**, 2745–2752 (1987).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank L. Lu and C. Glasser for technical assistance, J. Lidestri for maintenance of the X-ray laboratory at Columbia University, and W. N. Zagotta and S. Harrison for helpful discussions. Synchrotron diffraction data were collected at the x4a beamline at the National Synchrotron Light Source. The XL-I ultracentrifuge was obtained with funds provided by the NIH. This work was supported by the Klingenstein Foundation (E.G.), the National Alliance for Research on Schizophrenia and Depression (E.G.) and the NIH (E.G., M.L.M.). E.G. is also an assistant investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.G. (e-mail: jeg52@columbia.edu) or M.L.M. (e-mail: mlm@helix.nih.gov). The atomic coordinates for the crystal structures described here have been deposited at the Protein Data Bank under entry codes 1LB8 (L483Y-AMPA), 1LB9 (L483Y-DNQX), 1LBB (N754D-kainate) and 1LBC (N754S-Glu-CTZ).

Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis

M. B. Pepys*, J. Herbert*, W. L. Hutchinson*, G. A. Tennent*, H. J. Lachmann*, J. R. Gallimore*, L. B. Lovat*, T. Bartfai†‡, A. Alanine†, C. Hertel†, T. Hoffmann†, R. Jakob-Roetne†, R. D. Norcross†, J. A. Kemp†, K. Yamamura§, M. Suzuki§, G. W. Taylor||, S. Murray||, D. Thompson¶, A. Purvis¶, S. Kolstoe¶, S. P. Wood¶ & P. N. Hawkins*

* Centre for Amyloidosis and Acute Phase Proteins, Department of Medicine, Royal Free and University College Medical School, London NW3 2PF, UK

† Preclinical CNS Research, Pharmaceutical Division, F Hoffmann-La Roche Ltd, CH-4070 Basel, Switzerland

§ Institute of Molecular Embryology and Genetics, and Center for Animal Resources and Development, Kumamoto University, Japan

|| Division of Medicine, Imperial College School of Medicine, London W12 0NN, UK

¶ Division of Biochemistry and Molecular Biology, School of Biological Science, University of Southampton, Southampton SO16 7PX, UK

The normal plasma protein serum amyloid P component (SAP) binds to fibrils in all types of amyloid deposits, and contributes to the pathogenesis of amyloidosis. In order to intervene in this process we have developed a drug, *R*-1-[6-[*R*-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]pyrrolidine-2-carboxylic acid, that is a competitive inhibitor of SAP binding to amyloid fibrils. This palindromic compound also crosslinks and dimerizes SAP molecules, leading to their very rapid clearance by the liver, and thus produces a marked depletion of circulating human SAP. This mechanism of drug action potentially removes SAP from human amyloid deposits in the tissues and may provide a new therapeutic approach to both systemic amyloidosis and diseases associated with local amyloid, including Alzheimer's disease and type 2 diabetes.

Amyloidosis is a disorder of protein folding in which normally soluble globular proteins are deposited extracellularly in the tissues as abnormal insoluble cross- β -sheet fibrils, leading to tissue damage and disease^{1,2}. Treatments that reduce the supply of fibril precursor proteins can lead to regression of amyloid deposits^{3–6}, but such an approach is not possible in most forms of either acquired or hereditary amyloidosis, and these disorders are usually fatal. There is no treatment that specifically promotes regression of amyloid deposits, and systemic amyloidosis—in which the deposits are unequivocally the direct cause of disease—is responsible for about one per thousand of all deaths in developed countries. Amyloidosis is also a common and serious complication of long-term haemodialysis for end-stage renal failure. Furthermore, amyloid is always associated with the neurodegeneration of Alzheimer's disease and with the islet failure of type 2 diabetes. Although it is not known whether amyloid deposition itself causes the tissue dysfunction in these latter diseases, new treatments to promote removal of the damaging deposits are urgently required.

The normal non-fibrillar plasma glycoprotein serum amyloid P component, (SAP), a member of the pentraxin family, is universally present in amyloid deposits⁷. This reflects its specific calcium-dependent binding to motifs displayed by all types of amyloid fibrils. SAP itself is highly resistant to proteolysis, and binding of SAP to amyloid fibrils *in vitro* protects them from degradation by phagocytic cells and proteolytic enzymes⁸. Furthermore, SAP persists within human amyloid deposits for prolonged periods and is completely unmodified with respect to circulating SAP⁹. In mice and hamsters, plasma concentrations of SAP are closely related to amyloidogenesis⁷. We have therefore suggested that SAP may contribute to the failure to clear amyloid deposits *in vivo*, leading to tissue damage and disease. When we first identified a chemically defined ligand for SAP of low molecular mass—the cyclic pyruvate acetal of galactose, methyl 4,6-O-(1-carboxyethylidene)- β -D-galac-

topyranoside (MO β DG)—and showed that it could completely dissociate SAP from amyloid deposits *in vitro*, we proposed that such molecular dissection of the deposits might be an approach to therapy of amyloidosis¹⁰. Our recent finding that mice with targeted deletion of the SAP gene show retarded and reduced induction of experimental reactive systemic (AA type) amyloidosis confirmed

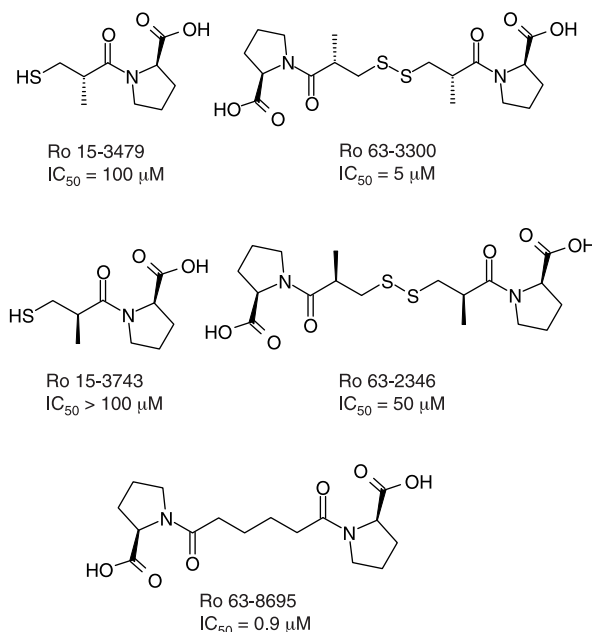


Figure 1 Inhibitors of SAP binding. The original screening hits were Ro 15-3479 and Ro 63-2346. Ro 63-3300 was synthesized and was more potent. The optimal product of the subsequent chemistry programme was Ro 63-8695 (*R*-1-[6-[*R*-2-carboxy-pyrrolidin-1-yl]-6-oxo-hexanoyl]pyrrolidine-2-carboxylic acid), which we abbreviate here as CPHPC. The IC₅₀ values shown are as measured in the primary screening assay.

‡ Present address: Department of Neuropharmacology, The Scripps Research Institute, La Jolla, California 92037, USA.

that SAP does indeed contribute to pathogenesis of amyloidosis¹¹ and is a valid therapeutic target¹².

Identification and development of inhibitors of SAP binding

A high throughput assay for inhibitors of SAP binding to Alzheimer's disease amyloid- β (A β) amyloid fibrils¹² immobilized in microtitre plates was used to screen the Roche compound library. Two hits were identified: Ro 15-3479, 1-[S-3-mercapto-2-methylpropionyl]-D-proline, an epimer of captopril, and Ro 63-2346, R-1-[R-3-[R-2-carboxy-pyrrolidin-1-yl]-2-methyl-3-oxo-propyl-disulphanyl]-2-methyl-propionyl-pyrrolidine-2-carboxylic acid, the oxidative disulphide dimer of the diastereomer 1-[R-3-mercapto-2-methylpropionyl]-D-proline (Ro 15-3743; Fig. 1). Captopril, 1-[S-3-mercapto-2-methylpropionyl]-L-proline, was inactive, demonstrating the key importance of the D-proline configuration. The dimer of the first hit was prepared—R-1-[S-3-[S-3-[R-2-carboxy-pyrrolidin-1-yl]-2-methyl-3-oxo-propyl-disulphanyl]-2-methyl-propionyl]-pyrrolidine-2-carboxylic acid, Ro 63-3300 (Fig. 1)—and was a more potent inhibitor of SAP binding than the monomeric thiol. The molecular mechanism responsible for this greater potency was suggested by our recent observation that the complex between SAP and dAMP (2'-deoxyadenosine 5'-phosphate) contained pairs of pentameric SAP molecules, interacting face to face as a result of base-stacking interactions between the bound dAMP molecules¹³. The pairing of SAP molecules bridged by the palindromic compound Ro 63-3300 would both enhance the avidity of the protein-ligand interaction and occlude the binding (B) face¹⁴ of the SAP molecule. Gel filtration experiments confirmed that pairs of SAP molecules were indeed formed in mixtures with Ro 63-3300 at appropriate molar ratios. Ro 63-3300 was selected as the lead compound for synthetic work to optimize the half-maximal inhibitory concentration (IC₅₀) for inhibition of SAP binding to amyloid fibrils *in vitro*, enhance the capacity to inhibit/dissociate SAP binding to amyloid deposits *in vivo*, and to minimize toxicity *in vivo*. *In vivo* evaluation of the best candidate compounds led to selection of R-1-[6-[R-2-carboxy-pyrrolidin-1-yl]-6-oxohexanoyl]pyrrolidine-2-carboxylic acid¹⁵ (Ro 63-8695), abbreviated to CPHPC, as the optimal compound for clinical proof of concept testing.

Molecular interaction between SAP and CPHPC

SAP is a pentamer with five identical non-covalently associated protomers each bearing a single calcium-dependent ligand-binding site on one planar face, the binding (B) face, of the molecule¹⁶. The palindromic structure of CPHPC, with two D-proline residues joined by an aliphatic linker chain, enables it, like Ro 63-3300, not only to block the ligand-binding sites on individual SAP protomers, but also to crosslink pairs of pentameric SAP molecules to form B-face to B-face decamers. Gel filtration analysis confirmed that in mixtures of isolated SAP with CPHPC, at ratios between equimolar and 100-fold molar excess of CPHPC (relative molecular

mass (M_r) of 340) over SAP protomers (M_r 25,462), all the SAP was decameric. However at 128-fold or greater molar excess of CPHPC, all the SAP was in its single pentameric form, presumably because each binding site was then occupied by a different individual CPHPC molecule, preventing crosslinking of pentamers. Addition of CPHPC to whole serum, at a concentration equimolar with SAP or greater, also produced stable decameric SAP. The multi-point binding of CPHPC between pairs of pentameric SAP molecules greatly increases the avidity of the interaction, and this was reflected in an apparent affinity constant of 10 nM, measured by isothermal calorimetry. This contrasted with affinity constants of just 50 μ M for MO β DG and 15 μ M for N-acetyl D-proline. Interestingly, the dissociation constant for binding of human SAP to isolated amyloid fibrils is also about 10 nM¹⁷.

The X-ray crystal structure of the SAP-CPHPC complex determined at 3.3 Å resolution (Table 1 and Fig. 2) confirmed that it was a decamer composed of two pentameric SAP molecules reversibly crosslinked by five CPHPC molecules. The CPHPC terminal carboxylates are bound in the calcium-dependent ligand-binding pockets of the SAP subunits and bridge across between the adjacent B faces of the two SAP molecules (Fig. 2). There are no close contacts between the two pentameric SAP molecules and they are separated by a clear solvent region. Further stabilization of the complex derives from the snug binding of the pyrrolidine ring of the drug into a hydrophobic pocket formed by Tyr 74, Leu 62 and Tyr 64 adjacent to the bound calcium ions. The 2.4 Å resolution structure of SAP co-crystallized with N-acetyl-D-proline shows a very close superposition with the drug head group in the CPHPC complex, with the proline side-chain pucker enabling packing against the Tyr 74 ring. The alkyl chain linking the two proline head groups of the drug is too long to fit the electron density in an extended conformation and adopts a kinked rotamer with eclipsed substituents about the C2-C3 bond. This displaces the two half-drug components (the halves of the palindromic CPHPC molecule) from a common long axis and facilitates approach of the head groups to the binding sites of two-fold axis related SAP protomers. Furthermore, the best fit to the electron density is achieved when both of the peptide bonds preceding the proline residues adopt the *cis*-configuration. On energetic grounds this *cis*-peptide bond probably occurs in about 25% of the drug head groups in solution, but the species carrying two *cis* proline head groups is probably much less abundant. The impact of this observation on drug potency is the subject of further studies.

Experimental studies of CPHPC *in vivo*

CPHPC was not metabolized in mice and was very rapidly excreted, predominantly in the urine, with a small amount in the bile. However, intraperitoneal or subcutaneous injection of CPHPC inhibited uptake of radiolabelled human SAP tracer into experimentally induced mouse AA amyloid deposits. Continuous infusion of CPHPC for 5 days accelerated whole-body clearance of radiolabelled human SAP tracer, with which the amyloid deposits had previously been loaded, and removed all the endogenous mouse SAP from the deposits (Fig. 3). Even 50 μ g per kg per day of CPHPC significantly dissociated human SAP from the deposits (not shown), but 1.5 mg per kg per day was required to dissociate any endogenous mouse SAP significantly, and there was a clear dose response effect up to 15 mg per kg per day, which removed all of the mouse SAP (Fig. 3). Injection of just 100 μ g CPHPC into mice transgenic for human SAP reduced circulating human SAP values by greater than 95% within 3 h, whereas the same dose of the monomeric thiol Ro 15-3479 (Fig. 1) had no effect at all on the plasma concentration of human SAP despite being able to inhibit uptake of radiolabelled human SAP into murine AA amyloid deposits *in vivo* (not shown). The capacity of the palindromic CPHPC molecule to crosslink pairs of SAP molecules is evidently critical for depletion of circulating SAP. Despite very limited oral bioavailability, administration of

Table 1 Data collection and refinement statistics for the SAP-CPHPC complex

Parameter	Value
Space group	$P4_32_12$
Unit cell (Å)	$a = b = 230.9, c = 281.4$
Resolution range (Å)	25–3.3
Measured reflections	423,005
Unique reflections	110,179
Multiplicity	3.8 (3.2)
Completeness (%)	96.7 (90.9)
R_{merge} (%)	8.2 (46.0)
$\langle I/\sigma \rangle$ (%) >2.0	74.0 (16.2)
Solvent content (%)	83.3
Model R_{factor} (%)	22.3
Model R_{free} (%)	22.9
r.m.s. bond lengths (Å)	0.010
r.m.s. bond angles (degree)	1.60

Values in parentheses denote the highest resolution shell statistics between 3.42–3.3 Å. r.m.s., root mean square. $\langle I/\sigma \rangle$, intensity/standard deviation of intensity.

CPHPC in the drinking water also rapidly induced sustained depletion of circulating human SAP in the transgenic mice (Fig. 3). However, none of these CPHPC treatments reduced the plasma concentration of mouse SAP in wild-type animals. The human SAP–CPHPC complex is evidently recognized as abnormal *in vivo* and is swiftly removed from the circulation, but mouse SAP–CPHPC complexes are apparently formed and/or cleared less efficiently *in vivo*.

Murine AA amyloidosis, induced experimentally using chronic inflammation to produce persistent high levels of serum amyloid A protein (SAA), the precursor of AA amyloid fibrils², is not suitable for rigorous studies of amyloid regression because, once initiated, amyloid deposition proceeds unpredictably throughout life. Nevertheless we investigated the effect of CPHPC on this model, but, to specifically target human SAP, we used SAP knockout mice transgenic for human SAP that express human SAP instead of mouse

SAP. Amyloid was induced with a single intravenous injection of amyloid-enhancing factor followed by four subcutaneous injections of casein over the next 7 days¹¹. Seven days later treatment was started in 21 mice with 0.25–1.0 mg ml^{−1} CPHPC in the drinking water, sufficient to reduce and maintain circulating human SAP values at 5 mg l^{−1} or less, while a control group ($n = 11$) received no treatment. After 40 days treatment, amyloid deposition was quantified by three independent observers (who were blind to the treatment regime) using Congo red histochemical staining¹¹. There was significantly less amyloid in the treated group ($P < 0.0001$, Wilcoxon Mann–Whitney test), and 5 of the 11 untreated controls had a substantial amyloid load compared with only 2 of the 21 CPHPC-treated mice ($P = 0.032$, Fisher's exact test), suggesting that treatment with CPHPC had reduced the amyloid burden.

Formal toxicological assessment of CPHPC in rats and dogs revealed no adverse effects during 28 days intravenous adminis-

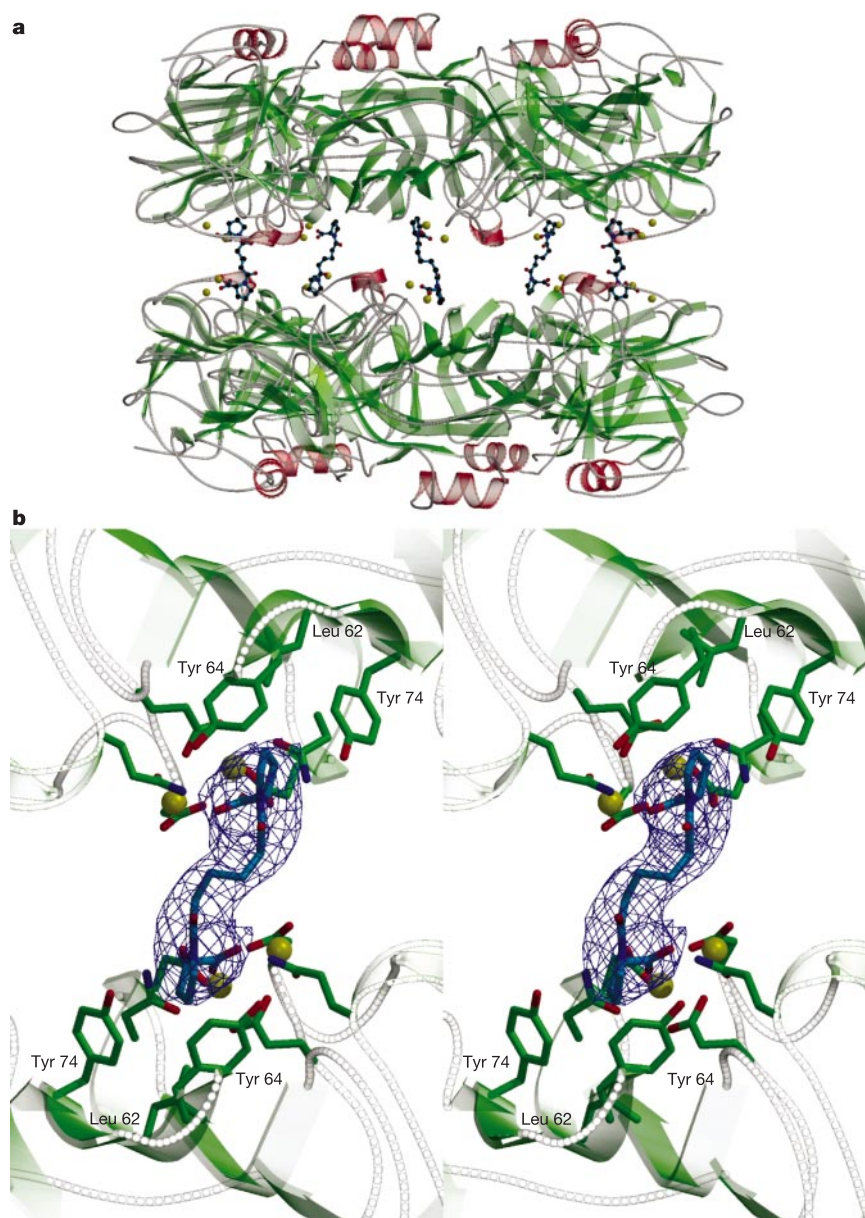


Figure 2 The structure of the complex of CPHPC with SAP. **a**, Two SAP pentamers crosslinked by means of their B faces by five molecules of CPHPC (blue), viewed perpendicular to the fivefold axis. A-face helices are shown in red. The two calcium ions bound to each SAP subunit are yellow. **b**, Stereo view of annealed omit map electron

density ($(F_o - F_c)$ contoured at 3σ) and the fitted CPHPC molecule showing the kinked alkyl linker and binding of the drug head groups into the double calcium site of two-fold axis related SAP subunits. Images were prepared with Bobscrip³⁹ and Raster3d⁴⁰.

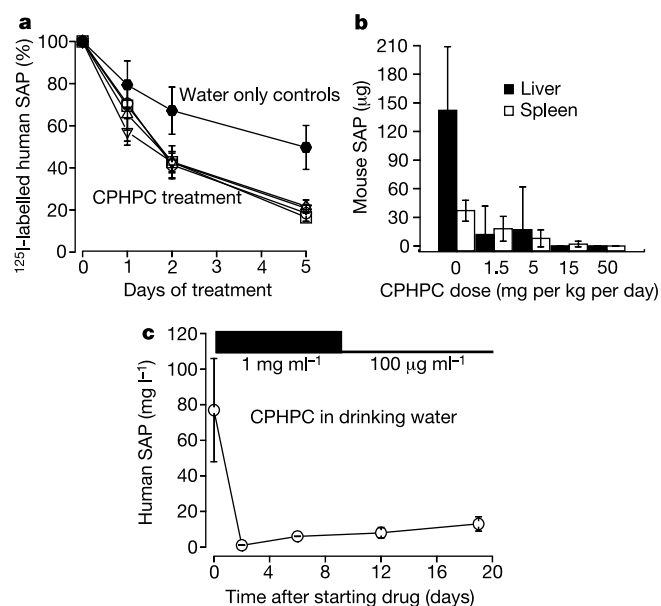


Figure 3 Effects of CPHPC on human and mouse SAP in mice *in vivo*. **a**, **b**, Groups of ten age-, sex- and weight-matched mice with experimentally induced reactive systemic AA amyloidosis were given a single loading dose of ¹²⁵I-labelled human SAP on day -1, and received implanted osmotic mini-pumps delivering the doses shown of CPHPC. Clearance from the amyloid deposits, catabolism and excretion of the human SAP tracer were monitored globally by whole-body counting of the mice (**a**). Each point represents the mean (s.d.) of all animals in each group; there was no significant difference between the groups receiving different doses of CPHPC: 1.5 mg per kg per day (squares); 5.0 mg per kg per day (circles); 15 mg per kg per day (triangles); 50 mg per kg per day (inverted triangles). The total amount of mouse SAP in the amyloidotic organs was determined after killing the animals on day 5 (**b**). **c**, Effect of administration of CPHPC in the drinking water (at the concentrations shown) on circulating human SAP values in human SAP transgenic mice. Mice of approximately 20 g body weight consume about 3 ml of water per day.

tration of up to 400 mg per kg per day. The drug was not metabolized and 95% of a single dose was excreted within 24 h. During prolonged administration, a mean of 99% (s.d. 1%) of the drug was recovered, 90% (s.d. 3%; range 85–93%) from renal excretion and 8% (s.d. 3%; range 6–13%) in the faeces. CPHPC did not inhibit the five main human cytochrome P450 isoenzymes and was, therefore, unlikely to affect the metabolism of other drugs. We thus proceeded to study directly the action of CPHPC in patients with amyloidosis.

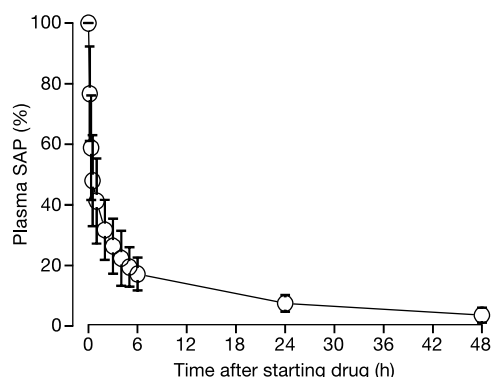


Figure 4 Effect of intravenous infusion of CPHPC on plasma SAP values in patients with systemic amyloidosis. Six patients with AL amyloidosis and one with AA type received doses of CPHPC of between 0.25–6.0 mg per kg per day for 48 h. The SAP concentration was measured immediately before and at the intervals shown during the infusion. Each point represents the mean (s.d.) of all patients.

Clinical studies of CPHPC in human amyloidosis

CPHPC was administered for 48 h by intravenous infusion to seven patients with systemic amyloidosis (6 monoclonal immunoglobulin light chain (AL) type and one AA type), and one patient with minor localized AL amyloidosis. There was rapid and consistent depletion of circulating SAP in all subjects—the SAP concentration started to fall when about 2 mg CPHPC had been given—and at all doses between 0.25–6 mg per kg per day, SAP was almost completely cleared from the plasma by the end of the infusion (Fig. 4). Circulating SAP values were already falling sharply when the ratio of plasma drug concentration to that of SAP pentamer approached equimolar levels. Gel filtration analysis confirmed that this SAP:CPHPC ratio is sufficient to generate some decameric SAP, but presumably with pairs of pentameric SAP molecules crosslinked by just two CPHPC molecules rather than by five as in the crystal structure (Fig. 2). These pairs of SAP molecules were evidently recognized as abnormal and were promptly cleared. In one further subject—a carrier of the amyloidogenic Ala 60 transthyretin variant but with no clinical amyloidosis—a bolus of 2 mg CPHPC followed by infusion at just 0.1 mg per kg per day produced slower and less complete SAP depletion.

The plasma SAP concentration returned to normal by 48 h after stopping the drug in the individuals with little or no amyloid, but in subjects with significant amyloidosis the SAP value remained low for prolonged periods of time. In the patient in the group with the heaviest whole-body amyloid load, the plasma SAP concentration was still below 25% of its initial value 20 days after the infusion. This indicates that most of the 50–100 mg of SAP newly produced per day¹⁸ was distributing into the amyloid deposits before becoming available to replete the plasma pool, and it is very strong indirect evidence that even the brief 48-h infusion of CPHPC had substantially depleted the amyloid-associated SAP pool.

Direct evidence for depletion of SAP from amyloid in the organs and for the mechanism of action of CPHPC was obtained by quantitative whole-body scintigraphy using ¹²³I-labelled SAP as a tracer. Each patient received a standard dose of ¹²³I-labelled SAP 24 h before the CPHPC infusion started, and was scanned immediately before treatment to provide a baseline image and values for localization of tracer to the amyloid deposits. Patients were then scanned at intervals thereafter, up to the end of the 48-h infusion. By 6 h after starting treatment the blood pool signal had virtually disappeared (Fig. 5), and there was marked accumulation of tracer in the liver, identifying this organ as the site to which CPHPC caused clearance of the circulating SAP. At the same time there was a

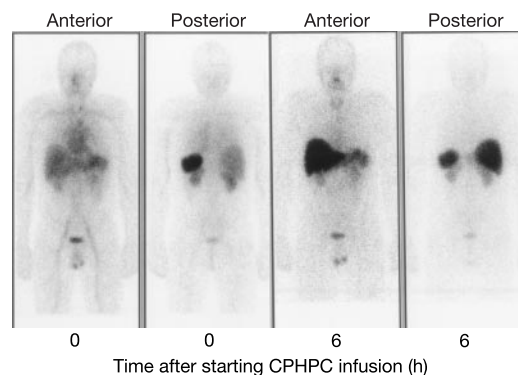


Figure 5 Whole-body ¹²³I-labelled SAP scintigraphy before and 6-h after starting infusion of CPHPC. The figure shows a patient with a modest load of AL amyloid in the spleen and kidneys, with notable blood pool background of tracer in the heart and circulation before administration of the drug (left panel). At 6 h the blood pool background is completely absent and the liver, which is the only site of catabolism of SAP *in vivo*, has taken up the tracer while the intensity of SAP signal from the amyloidotic spleen and kidneys is significantly reduced.

marked decrease in the retention of tracer in amyloid deposits elsewhere, exemplified by the spleen, in contrast to the usual situation in untreated (control) amyloidosis patients (Table 2).

We have previously demonstrated in mice that the liver, and specifically the hepatocyte, is the only significant site of clearance and catabolism of both mouse and human SAP *in vivo*¹⁹. Furthermore, asialo human SAP is instantly cleared by the liver in humans by means of the hepatocyte asialoglycoprotein receptor, and we have imaged this process using ¹²⁵I-labelled asialo SAP⁹. CPHPC apparently triggers similarly rapid hepatic uptake of SAP *in vivo*, leading to virtually total depletion of circulating SAP. This promotes redistribution of SAP from the tissues to the plasma, supplementing the effect of CPHPC as a competitive inhibitor of SAP binding to amyloid fibrils, and enhancing the efficiency of removal of SAP from amyloid deposits. Importantly from a clinical viewpoint, subcutaneous injection of as little as 0.25 mg per kg per day of CPHPC, given in doses every 12 hours, had the same effect on plasma SAP as intravenous infusion of the drug.

In an open label study we have now infused CPHPC intravenously, or injected it subcutaneously, at 30–40 mg per day (mean, 38.7 mg) continuously for periods of 1.2–9.5 months (mean, 2.6 months) in 19 systemic amyloidosis patients aged 43–67 years (mean 56 ± 7 years). There were 12 patients with AL amyloidosis, two with hereditary transthyretin type, one with AA type, and one each with hereditary apolipoprotein AI, fibrinogen A α-chain, gelsolin and lysozyme amyloidosis². Plasma SAP values were reduced to a mean of 5 ± 2.5% of pre-treatment levels in all patients throughout treatment. There were no adverse clinical effects, either locally from twice daily subcutaneous injections, or systemically, and no biochemical, haematological, immunological or organ function abnormalities were detected that could be attributed to CPHPC. Although many of these patients had end-stage disease and were offered CPHPC because they had either already failed to respond to all other possible treatments, or no other treatment exists for their type of amyloidosis, most of them remained clinically stable throughout treatment. One individual with rapidly progressive, terminal, AL amyloidosis (affecting mainly the heart, kidneys, spleen and autonomic nervous system) who had not responded to maximal cytotoxic chemotherapy, lived for 6 months after starting treatment with CPHPC. At autopsy, the organs still contained substantial amyloid deposits but their SAP content was markedly reduced, to about 15% of the amount typically associated with comparable amyloid loads. Prolonged treatment with CPHPC is thus likely to be required to completely clear SAP from major amyloid deposits and to thereby maximally enhance amyloid regression.

Discussion

We report here the first example, to our knowledge, of a low molecular mass drug that profoundly depletes a specific plasma protein from the circulation and the tissues. The physiological mechanisms for recognizing extremely subtle, abnormal molecular conformations in plasma proteins are exquisitely sensitive and provide highly effective clearance from the circulation. The present

compound, CPHPC, takes advantage of this phenomenon and specifically targets human SAP to produce a 'knockout' for therapeutic purposes²⁰.

SAP in amyloid deposits derives entirely from circulating SAP, which is produced, and eventually cleared and catabolized, exclusively in the liver⁷. The plasma and amyloid pools of SAP are in free dynamic equilibrium, and our original goal was to produce a low molecular mass compound that would remove SAP from amyloid deposits *in vivo* by inhibiting binding of SAP to amyloid fibrils and dissociating SAP that was already bound. Indeed, CPHPC and the related molecules from which it was developed (Ro 15-3479 and Ro 63-3300, Fig. 1) do precisely that with respect to mouse SAP in murine systemic amyloidosis. However, human SAP binds more avidly than mouse SAP to all their known ligands including amyloid fibrils²¹, and binding of the palindromic CPHPC molecule by human SAP has an additional and powerful effect *in vivo*. CPHPC crosslinks pairs of native pentameric human SAP molecules, creating a decameric assembly that is swiftly cleared from the circulation by the liver. The resulting profound depletion of plasma SAP markedly shifts the distribution of SAP between the plasma and amyloid pools, with SAP flowing from the tissues into the circulation where it is immediately targeted by the drug for clearance in the liver. Coupled with direct inhibition of SAP binding by CPHPC, this leads to rapid and extensive removal of SAP from the amyloid deposits.

Binding of SAP stabilizes amyloid fibrils *in vitro* and protects them from proteolytic degradation. We therefore hope that efficient removal of SAP *in vivo* will reduce the stability of amyloid deposits and promote their regression. The absence of SAP may also retard new amyloid deposition, as it does in SAP-deficient mice¹¹. We are currently testing this hypothesis by long-term treatment of systemic amyloidosis patients with CPHPC. Elimination of new fibril formation by removal of the supply of fibril precursor proteins is associated with regression of deposits and with clinical benefit in all forms of amyloidosis in which it has been achieved^{3–6}, indicating that clearance of human amyloid deposits can occur *in vivo*. Clearance of experimental amyloid deposits has been reported in transgenic mice expressing human Aβ, the amyloid fibril protein of Alzheimer's disease, after production or administration of anti-Aβ antibodies^{22,23}, and inhibition or reversal of amyloid fibrillogenesis are also being investigated. However, CPHPC has potential advantages over other therapeutic approaches and could also be combined with them. It is an easily administered, non-toxic, well tolerated and highly potent low molecular mass compound. Importantly, the drug may not need to gain access to the tissues as it exerts its desired biochemical effect within the circulation. Thus if CPHPC does indeed promote amyloid regression, it should be readily applicable not only in systemic amyloidosis where disease is unequivocally caused by the deposits, but also in both Alzheimer's disease and type 2 diabetes, in which local amyloid deposits are implicated in pathogenesis. □

Table 2 CPHPC rapidly clears circulating SAP to the liver and depletes SAP from visceral amyloid deposits

Treatment	¹²⁵ I-labelled SAP retention (%)					
	Liver			Spleen		
	Time after tracer injection (h)	Number of patients		Time after tracer injection (h)	Number of patients	
	24 h			24 h		
None	100	12		100	14	
CPHPC	100	7		100	7	
P-value		P < 0.0001			P = 0.008	

Values are mean values with s.d. in parentheses. All patients had systemic AL amyloidosis and received a standard intravenous tracer dose of ¹²⁵I-labelled SAP at time zero. After whole-body quantitative scintigraphic imaging at 24 h, uptake in liver and spleen were taken as 100% for each individual. Intravenous infusion of CPHPC was then started and scintigraphy with organ counting was repeated at 48 h. The controls received no treatment. Significance of differences between the two groups was evaluated by a *t*-test.

Methods

Protein studies

Human and mouse SAP were isolated, purified and assayed as previously described^{17,24–26}. Gel filtration analysis to determine the molecular assembly of SAP–CPHPC complexes was performed as reported elsewhere²⁷. Isothermal calorimetry to measure heat of binding between SAP and its ligands, and calculation of affinity constants were performed as previously reported²⁸ using a VP–ITC system from MicroCal Inc. Forty injections of 5 μ l of ligand were titrated at 5.4 min intervals into SAP, in solution in 10 mM Tris-buffered 0.14 M NaCl containing 10 mM EDTA and 200 mM CaCl₂, pH 8.0, within a 1.3-ml sample cell. Ligand concentration was varied between 1.6 mM and 800 μ M, and protein between 40 μ M and 100 μ M, so that a heat change of approximately 1 μ cal was observed on initial injections.

X-ray crystal structure

Crystals of the SAP–CPHPC complex were grown by hanging-drop vapour diffusion at pH 7.6 and 4 °C, using polyethylene glycol monomethyl ether 550 as the precipitant in the presence of a tenfold molar excess of CPHPC. The crystals were tetragonal, space group *P*4₃2₁2, with unit cell dimensions *a* = *b* = 230.9 Å and *c* = 281.4 Å. X-ray diffraction data were collected from a single crystal at 100 K to a resolution of 3.2 Å on beamline ID14-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, and were processed using DENZO and SCALEPACK²⁹. Molecular replacement was performed in MOLREP³⁰ using a previously derived SAP pentamer¹⁶ as the search model. There were two SAP pentamers in the asymmetric unit, with an unusually high solvent content of 83%. The structure was refined using CNS³¹ with tight tenfold NCS restraints between the protomers. During the course of the refinement the stereochemical properties were checked using WHAT IF³² and PROCHECK³³, with the final model showing no residues within the disallowed regions of the Ramachandran plot. A summary of the data collection and refinement statistics is shown in Table 1.

In vivo mouse studies

Reactive systemic AA amyloidosis was induced in CBA mice by repeated subcutaneous injections of casein, and studies of the localization, retention and turnover of radiolabelled human and mouse SAP, and of amyloid load, were conducted precisely as described previously^{11,21}. There was complete concordance between the rank scores, 0–5, assigned by different observers to the amount of amyloid demonstrable by Congo red staining in the spleen and liver. The sum of these scores for each animal, as well as simply the presence or absence of substantial deposits, were analysed by appropriate nonparametric statistical tests. Pure line C57BL/6 mice transgenic for human SAP were bred in London from embryos of the strain previously reported³⁴.

In vivo human studies

SAP scintigraphy studies, including whole-body imaging, region of interest counting of specific organs, and grading of whole-body amyloid load, were performed as described elsewhere^{35–38}. Plasma concentration of CPHPC was determined by Multiple Reaction Monitoring on a Quattro II mass spectrometer (Micromass) after protein precipitation, Oasis HLB extraction (Waters Ltd), dimethylation with diazomethane and high-performance liquid chromatography (10 cm $\times$ 2 mm; RPB) (acetonitrile:water:formic acid, 40:60:0.1 v/v/v, 0.14 ml min^{−1}). Transitions were mass/charge ratio *m/z* 369.4 to 240.2 (CPHPC) and 377.4 to 248.2 (²H₈-CPHPC).

Received 22 January; accepted 10 April 2002.

- Pepys, M. B. in *Protein Misfolding and Disease* (eds Dobson, C. M., Ellis, R. J. & Fersht, A. R.) *Phil. Trans. R. Soc. Lond. B* **356**, 203–211 (The Royal Society, London, 2001).
- Pepys, M. B. & Hawkins, P. N. in *Santer's Immunologic Diseases* (eds Austen, K. F., Frank, M. M., Atkinson, J. P. & Cantor, H.) 401–412 (Lippincott Williams, Philadelphia, 2001).
- Holmgren, G. *et al.* Clinical improvement and amyloid regression after liver transplantation in hereditary transthyretin amyloidosis. *Lancet* **341**, 1113–1116 (1993).
- Gillmore, J. D., Hawkins, P. N. & Pepys, M. B. Amyloidosis: a review of recent diagnostic and therapeutic developments. *Brit. J. Haematol.* **99**, 245–256 (1997).
- Gillmore, J. D. *et al.* Curative hepatorenal transplantation in systemic amyloidosis caused by the Glu526Val fibrinogen α -chain variant in an English family. *Q. J. Med.* **93**, 269–275 (2000).
- Gillmore, J. D., Lovat, L. B., Persey, M. R., Pepys, M. B. & Hawkins, P. N. Amyloid load and clinical outcome in AA amyloidosis in relation to circulating concentration of serum amyloid A protein. *Lancet* **358**, 24–29 (2001).
- Pepys, M. B. *et al.* Amyloid P component. A critical review. *Int. J. Exp. Clin. Invest.* **4**, 274–295 (1997).
- Tennent, G. A., Lovat, L. B. & Pepys, M. B. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer's disease and systemic amyloidosis. *Proc. Natl Acad. Sci. USA* **92**, 4299–4303 (1995).
- Pepys, M. B. *et al.* Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc. Natl Acad. Sci. USA* **91**, 5602–5606 (1994).
- Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L. & Pepys, M. B. Specific chemical dissociation of fibrillar and non-fibrillar components of amyloid deposits. *Lancet* **ii**, 376–378 (1984).
- Botto, M. *et al.* Amyloid deposition is delayed in mice with targeted deletion of the serum amyloid P component gene. *Nature Med.* **3**, 855–859 (1997).
- Pepys, M. B. & Blundell, T. L. Screening assays to identify therapeutic agents for amyloidosis. US patent 6126918 (2000).
- Hohenester, E., Hutchinson, W. L., Pepys, M. B. & Wood, S. P. Crystal structure of a decameric complex of human serum amyloid P component with bound dAMP. *J. Mol. Biol.* **269**, 570–578 (1997).
- Ashton, A. W., Boehm, M. K., Gallimore, J. R., Pepys, M. B. & Perkins, S. J. Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J. Mol. Biol.* **272**, 408–422 (1997).

- Hertel, C., Norcross, R. D., Jakob-Roetne, R. & Hoffmann, T. D-proline derivatives. US patent 6262089 (2001).
- Emsley, J. *et al.* Structure of pentameric human serum amyloid P component. *Nature* **367**, 338–345 (1994).
- Hawkins, P. N., Tennent, G. A., Woo, P. & Pepys, M. B. Studies *in vivo* and *in vitro* of serum amyloid P component in normals and in a patient with AA amyloidosis. *Clin. Exp. Immunol.* **84**, 308–316 (1991).
- Hawkins, P. N., Wootton, R. & Pepys, M. B. Metabolic studies of radioiodinated serum amyloid P component in normal subjects and patients with systemic amyloidosis. *J. Clin. Invest.* **86**, 1862–1869 (1990).
- Hutchinson, W. L., Noble, G. E., Hawkins, P. N. & Pepys, M. B. The pentraxins, C-reaction protein and serum amyloid P component, are cleared and catabolized by hepatocytes *in vivo*. *J. Clin. Invest.* **94**, 1390–1396 (1994).
- Pepys, M. B. Therapeutic protein depletion. GB Patent application No. 0119370.5 (filed 2001).
- Hawkins, P. N., Myers, M. J., Epenetos, A. A., Caspi, D. & Pepys, M. B. Specific localization and imaging of amyloid deposits *in vivo* using ¹²⁵I-labeled serum amyloid P component. *J. Exp. Med.* **167**, 903–913 (1988).
- Schenk, D. *et al.* Immunization with amyloid- β attenuates Alzheimer-disease-like pathology in the PDAPP mouse. *Nature* **400**, 173–177 (1999).
- Bard, F. *et al.* Peripherally administered antibodies against amyloid β -peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
- Pepys, M. B. Isolation of serum amyloid P component (protein SAP) in the mouse. *Immunology* **37**, 637–641 (1979).
- Pepys, M. B., Baltz, M., Gomer, K., Davies, A. J. S. & Doenhoff, M. Serum amyloid P-component is an acute-phase reactant in the mouse. *Nature* **278**, 259–261 (1979).
- Nelson, S. R. *et al.* Serum amyloid P component in chronic renal failure and dialysis. *Clin. Chim. Acta* **200**, 191–200 (1991).
- Hutchinson, W. L., Hohenester, E. & Pepys, M. B. Human serum amyloid P component is a single uncomplexed pentamer in whole serum. *Mol. Med.* **6**, 482–493 (2000).
- Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137 (1989).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56 (1990).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Iwanaga, T. *et al.* Liver-specific and high-level expression of human serum amyloid P component gene in transgenic mice. *Dev. Genet.* **10**, 365–371 (1989).
- Hawkins, P. N., Myers, M. J., Lavender, J. P. & Pepys, M. B. Diagnostic radionuclide imaging of amyloid: biological targeting by circulating human serum amyloid P component. *Lancet* **i**, 1413–1418 (1988).
- Hawkins, P. N., Lavender, J. P. & Pepys, M. B. Evaluation of systemic amyloidosis by scintigraphy with ¹²⁵I-labeled serum amyloid P component. *N. Engl. J. Med.* **323**, 508–513 (1990).
- Jager, P. L. *et al.* Kinetic studies with iodine-123-labeled serum amyloid P component in patients with systemic AA and AL amyloidosis and assessment of clinical value. *J. Nucl. Med.* **39**, 699–706 (1998).
- Rydh, A. L. Serum amyloid P component scintigraphy in familial amyloid polyneuropathy: regression of visceral amyloid following liver transplantation. *Eur. J. Nucl. Med.* **25**, 709–713 (1998).
- Esouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics Mod.* **15**, 132–134 (1997).
- Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Acknowledgements

This work was supported by grants from the Medical Research Council and The Wellcome Trust to M.B.P., P.N.H. and S.P.W., by National Health Service R & D Funds, and by many donations of patients, relatives and friends to the Amyloidosis Research Fund. M.B.P. conceived the idea of inhibition of SAP binding for therapy of amyloidosis, devised the screening assay, originated and directed the project and wrote the paper; J.H., W.L.H., G.A.T., J.R.G. and L.B.L. performed the experimental work and protein assays in the Pepys laboratory; T.B. and J.A.K. supervised the project within Roche; R.J.-R., R.D.N., A.A., C.H. and T.H. designed and synthesized the inhibitor molecules and evaluated them *in vitro*; K.Y. and M.S. provided the human SAP transgenic mouse line; G.W.T. and S.M. performed the *ex vivo* drug assays; S.P.W., D.T., S.K. and A.P. crystallized the SAP–drug complex and solved its structure; P.N.H. contributed to all aspects of the work in the Pepys laboratory and devised and supervised the clinical study, which was conducted by H.J.L. We thank R. Zell, C. Jenny (kilolab), A. FerRo, P. Keller (galenics), R. Strobel, E. Gocke, B. Schläppli (toxicology), D. Schwab, B. Levett-Traflet, J. Huwyler, S. Poli (pharmacokinetics), P. Williams, U. Widmer, Q. Branca & G. Adam for their contributions, S. Madhoo and D. Gopaul for care of the patients and the SAP scintigraphy studies, M. Botto for help in breeding transgenic mice, and E. Jones for preparation of the manuscript.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to M.B.P. (e-mail: m.pepys@rfc.ucl.ac.uk).

A reconstruction of the initial conditions of the Universe by optimal mass transportation

Uriel Frisch*, Sabino Matarrese†, Roya Mohayaee‡* & Andrei Sobolevski§*

* CNRS, UMR 6529, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France

† Dipartimento di Fisica "G. Galilei" and INFN, Sezione di Padova, via Marzolo 8, 35131-Padova, Italy

‡ Dipartimento di Fisica, Università Degli Studi di Roma "La Sapienza", P. le A. Moro 5, 00185-Roma, Italy

§ Department of Physics, M V Lomonossov University, 119899-Moscow, Russia

Reconstructing the density fluctuations in the early Universe that evolved into the distribution of galaxies we see today is a challenge to modern cosmology¹. An accurate reconstruction would allow us to test cosmological models by simulating the evolution starting from the reconstructed primordial state and comparing it to observations. Several reconstruction techniques have been proposed^{2–9}, but they all suffer from lack of uniqueness because the velocities needed to produce a unique reconstruction usually are not known. Here we show that reconstruction can be reduced to a well-determined problem of optimization, and present a specific algorithm that provides excellent agreement when tested against data from *N*-body simulations. By applying our algorithm to the redshift surveys now under way¹⁰, we will be able to recover reliably the properties of the primeval fluctuation field of the local Universe, and to determine accurately the peculiar velocities (deviations from the Hubble expansion) and the true positions of many more galaxies than is feasible by any other method.

Starting from the available data on the galaxy distribution, can we trace back in time and map to its initial locations the highly structured distribution of mass in the Universe (Fig. 1)? Here we show that, with a suitable hypothesis, the knowledge of both the present non-uniform distribution of mass and of its primordial quasi-uniform distribution uniquely determines the inverse lagrangian map, defined as the transformation from present (eulerian) positions *x* to the respective initial (lagrangian) positions *q*.

We first consider the direct lagrangian map *q* ↦ *x*, which can be approximately written in terms of a potential as *x* = ∇*q*Φ(*q*) at those scales where nonlinearity stays moderate¹¹. This is supported by numerical *N*-body simulations showing good agreement with a very simple potential approximation, due to Zel'dovich¹², which assumes that the particles move on straight trajectories. Even better agreement is obtained with a refinement, the second-order lagrangian perturbation method^{13–16}, also known to be potential.

In our 'reconstruction hypothesis', we furthermore assume the convexity of the potential Φ(*q*), a consequence of which is the absence of multi-streaming: for almost any eulerian position, there is a single lagrangian antecedent. As is well-known, the Zel'dovich approximation leads to caustics and to multi-streaming. This can be overcome in various ways, for example by a modification known as the adhesion model, an equation of viscous pressureless gas dynamics^{17,18}. The latter, which leads to shocks rather than caustics, is known to have a convex potential¹⁹ and to be in better agreement with *N*-body simulations. Suppression or reduction of multi-streaming requires a mechanism of momentum exchange, such as viscosity, between neighbouring streams having different velocities. This is a common phenomenon in ordinary fluids, such as the flow of air or water in our natural environment. Dark matter is, however, essentially collisionless, and the usual mechanism for generating

viscosity (discovered by Maxwell) does not operate, so that a non-collisional mechanism involving a small-scale gravitational instability must be invoked.

Our reconstruction hypothesis implies that the initial positions can be obtained from the present ones by another gradient map: *q* = ∇*x*Θ(*x*), where Θ is a convex potential related to Φ by a Legendre–Fenchel transform (see Methods). We denote by ρ₀ the initial mass density (which can be treated as uniform) and by ρ(*x*) the final one. Mass conservation implies ρ₀ d³*q* = ρ(*x*) d³*x*. Thus, the ratio of final to initial density is the jacobian of the inverse lagrangian map. This can be written as the following Monge–Ampère equation²⁰ for the unknown potential Θ:

$$\det(\nabla_{x_i} \nabla_{x_j} \Theta(\mathbf{x})) = \rho(\mathbf{x})/\rho_0 \quad (1)$$

where 'det' stands for determinant.

We emphasize that no information about the dynamics of matter other than the reconstruction hypothesis is needed for our method, whose degree of success depends crucially on how well this hypothesis is satisfied. Exact reconstruction is obtained, for example, for the Zel'dovich approximation (before particle trajectories cross) and for adhesion-model dynamics (at arbitrary times).

We note that our Monge–Ampère equation for self-gravitating matter may be viewed as a nonlinear generalization of a Poisson equation (used for reconstruction in ref. 4), to which it reduces if particles have moved very little from their initial positions.

It has been discovered recently that the map generated by the solution to the Monge–Ampère equation (1) is the (unique) solution to an optimization problem²¹ (see also refs 22 and 23).

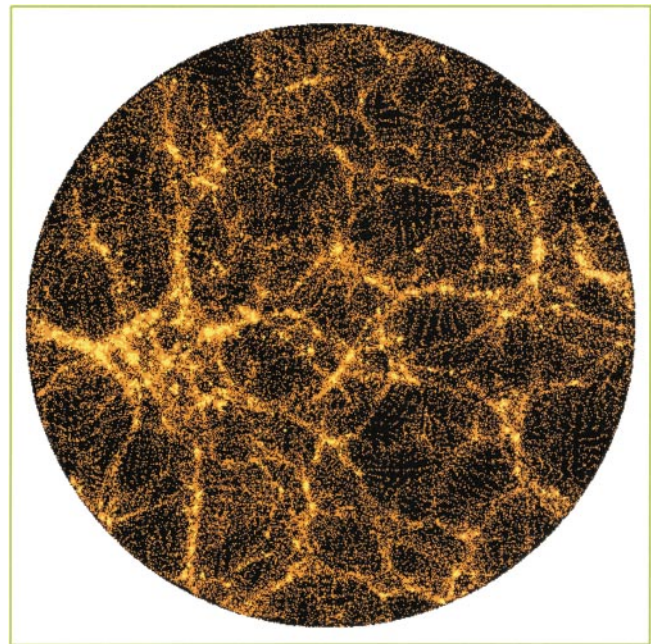


Figure 1 *N*-body simulation output (present epoch) used for testing our reconstruction method. In the standard model of structure formation, the distribution of matter in the Universe is believed to have emerged from a very smooth initial state: tiny irregularities of the gravitational potential, which we can still observe as temperature fluctuations of the cosmic microwave background, gave rise to density fluctuations, which grew under their self-gravity, developing a rich and coherent pattern of structures. Most of the mass is in the form of cold dark matter; the luminous matter (galaxies) can be assumed to trace—up to some form of bias—the underlying dark matter. Shown here is a projection onto the *x*–*y* plane of a 10% vertical slice of the simulation box of size 200 *h*^{−1} Mpc. The model, ΛCDM, uses cold dark matter with cosmological constant and the following parameters: Hubble constant *h* = 0.65, density parameters Ω_Λ = 0.7 and Ω_m = 0.3, normalization factor σ₈ = 0.99. Points are highlighted in yellow when reconstruction fails by more than 6 *h*^{−1} Mpc, which happens mostly in high-density regions.

This is the ‘mass transportation’ problem of Monge and Kantorovich^{24,25}, in which one seeks the map $\mathbf{x} \mapsto \mathbf{q}$ that minimizes the quadratic ‘cost’ function:

$$I = \int \rho_0 |\mathbf{x} - \mathbf{q}|^2 d^3 q = \int \rho(\mathbf{x}) |\mathbf{x} - \mathbf{q}|^2 d^3 x \quad (2)$$

Note that $\mathbf{x} = \mathbf{q}$ is forbidden: as the initial and final density fields ρ_0 and $\rho(\mathbf{x})$ are prescribed, there is a constraint on the jacobian of the map (see Methods).

Next, we take into account that information on the mass distribution is provided in the form of N discrete particles both in simulations and when handling observational data from galaxy surveys. The cost minimization then becomes what is known in optimization theory as the assignment problem: find the unique one-to-one pairing of a set of N initial points $\mathbf{q}_i$ and N final points $\mathbf{x}_i$ that minimizes $I_{\text{discr}} = \sum_{i=1}^N |\mathbf{x}_i - \mathbf{q}_{j(i)}|^2$. An immediate consequence is that, for any subset of k pairs of initial and final points ($2 \leq k \leq N$), the contribution of these points to the cost function should not decrease under arbitrary permutations of initial points. This property is known to be equivalent to having a lagrangian map that is the gradient of a convex function²⁶.

If we restrict ourselves to interchanging just pairs ($k = 2$), the map is said to be monotonic, a condition not equivalent to minimization of the cost function (except in one dimension). In ref. 5, a method of reconstruction called the path interchange Zel’dovich approximation (PIZA) is introduced, which uses the same quadratic cost function (obtained by applying a minimum-action argument within the framework of the Zel’dovich approximation). In PIZA, a randomly chosen tentative correspondence between initial and final points is successively improved by swapping pairs of initial particles whenever this decreases the cost

function. Eventually, a monotonic map is obtained that usually does not minimize the cost. This explains the non-uniqueness of PIZA reconstruction (also noticed in ref. 8).

There are, however, known deterministic strategies for the assignment problem that give the correct unique solution; their complexity (dependence on N of the number of operations needed) is close to N^3 for arbitrary cost functions, but can be sharply reduced when the cost function is quadratic. Combining the organization of data taken from Hénon’s mechanical analogue machine (see ref. 27, and <http://www.obs-nice.fr/etc7/henon.pdf>) for solving the assignment problem with the dual simplex method of Balinski²⁸, we have designed an algorithm that gives the optimal assignment for about 20,000 particles in a few hours of CPU time on a fast Alpha machine. For historical reasons, we call our approach Monge–Ampère–Kantorovich or MAK (see Methods). Details of the algorithms will be given elsewhere; we merely note that, when working with the catalogues of several hundred thousand galaxies that are expected within a few years, a direct application of the assignment algorithm in its present state could require unreasonable computational resources. A mixed strategy can however be used, in which the assignment problem is solved on a coarse grid while, on smaller scales, the Monge–Ampère equation (1) is solved by a relaxation technique (adapted from ref. 23).

We tested the MAK reconstruction on data obtained by a cosmological N -body simulation with 128^3 particles, using the HYDRA code²⁹ (Fig. 1). Reconstruction was performed on three 32^3 grids with (comoving) meshes given by $\Delta x = 6.25 h^{-1} \text{ Mpc}$, $\Delta x/2$ and $\Delta x/4$, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. In comoving coordinates, the typical displacement of our mass elements over one Hubble time is about $10 h^{-1} \text{ Mpc}$. We discarded those points that, at the end of the simulation (present epoch), were not within a sphere containing about 20,000 points, a number comparable to that of currently available all-sky galaxy redshift catalogues. As the simulation

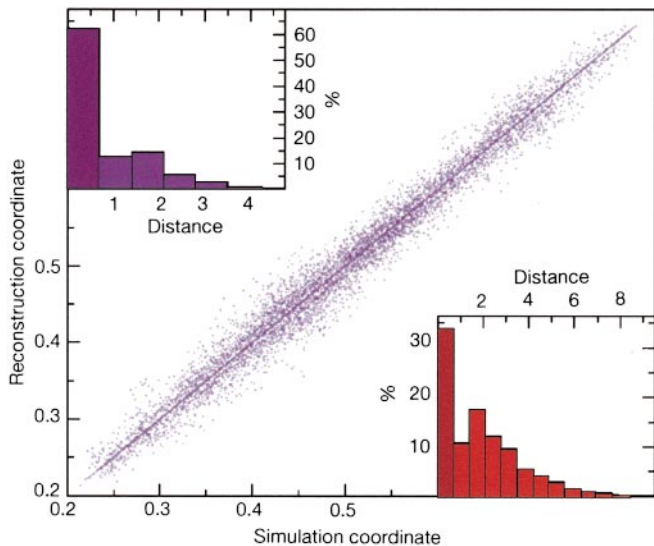


Figure 2 Tests of MAK reconstructions of the lagrangian positions, using the data shown in Fig. 1. The dots near the diagonal are a scatter plot of reconstructed initial points versus simulation initial points for the coarsest $6.25 h^{-1} \text{ Mpc}$ grid with 17,178 points. The scatter diagram uses a ‘quasi-periodic projection’ coordinate $\tilde{q} \equiv (q_1 + q_2\sqrt{2} + q_3\sqrt{3})/(1 + \sqrt{2} + \sqrt{3})$, which guarantees a one-to-one correspondence between $\tilde{q}$ -values and points on the regular lagrangian grid. The upper left inset is a histogram (by percentage) of distances in reconstruction mesh units between such points; the first bin (whose width was taken to be slightly less than one mesh) corresponds to perfect reconstruction (thereby allowing a good determination of the peculiar velocities of galaxies); the lower right inset is a similar histogram for reconstruction on a finer $3.12 h^{-1} \text{ Mpc}$ grid using 19,187 points. With the $6.25 h^{-1} \text{ Mpc}$ grid, 62% of the 17,178 points are assigned perfectly and about 75% are within not more than one mesh. With the $3.12 h^{-1} \text{ Mpc}$ grid, we have 34% of exact reconstruction out of 19,187 points. On further refinement of the mesh by a factor of two, this degrades to 14%.

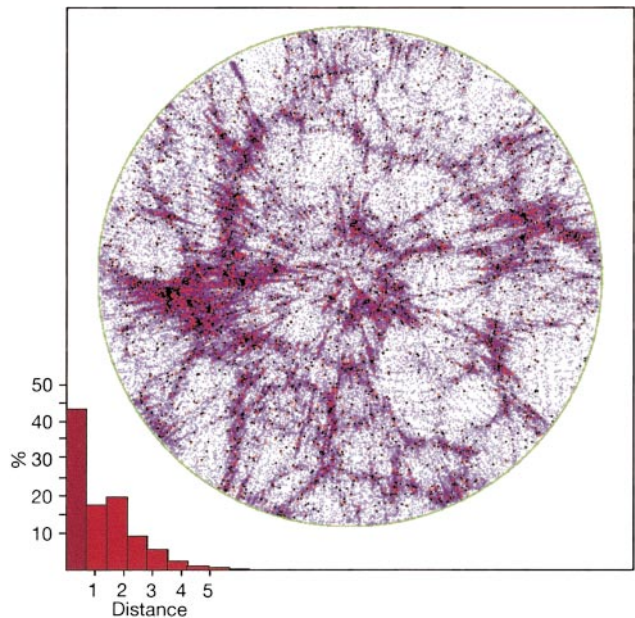


Figure 3 Reconstruction test in redshift space with the same data as for the real-space reconstruction tested in the upper left histogram of Fig. 2. The circular redshift map (violet points) corresponds to the same real-space slice as displayed in Fig. 1. The observer is taken to be at the centre of the simulation box. Points used for reconstruction within the displayed slice are highlighted in red. Reconstruction is performed by the MAK algorithm with a different cost function, obtained (as in ref. 8) by assuming that the peculiar velocities $\mathbf{v}$ can be estimated by the Zel’dovich approximation: $\mathbf{v} = f(\mathbf{x} - \mathbf{q})$, where $f \approx \Omega_m^{0.6} \approx 0.49$. Note that we now have 43% of exactly reconstructed points, out of the 60% which are within not more than $6.25 h^{-1} \text{ Mpc}$ from their correct positions.

assumes periodic boundary conditions, we also took into account periodicity when calculating the distance between pairs of points. The MAK reconstructions were used to generate a scatter diagram and various histograms allowing comparisons of simulation and reconstructed lagrangian points (Fig. 2). The results demonstrate the essentially potential character of the lagrangian map above $\sim 6 h^{-1}$ Mpc (within the given cosmological model).

We also performed PIZA reconstructions on the coarsest grid, and obtained typically 30–40% exactly reconstructed points, but severe non-uniqueness: for two different seeds of the random generator, only about half of the exactly reconstructed points were the same.

When reconstructing from observational data, in redshift space (Fig. 3), the galaxies appear displaced radially (as seen by the observer) by an amount proportional to the radial component of the peculiar velocity. We thus performed another reconstruction, with an accordingly modified cost function, that led to somewhat degraded results (Fig. 3) but at the same time provided an approximate determination of peculiar velocities. More-accurate determination of peculiar velocities can be done using second-order lagrangian perturbation theory. The effect of the catalogue selection function can be handled by standard techniques; for instance, one can assign each galaxy a 'mass' inversely proportional to the catalogue selection function^{8,9}.

What is the smallest length scale at which an optimization algorithm such as MAK can be expected to give a unique and reliable reconstruction? The key ingredient here is the simultaneous knowledge of the initial and present mass density fields. MAK-type reconstruction (with a suitable cost based on the equation of a self-gravitating fluid) should therefore be possible down to scales comparable to the thickness of collapsed structures, below which the hydrodynamical description ceases to be meaningful.

The fact that MAK guarantees a unique solution, and that our present reconstruction hypothesis proved to be very faithful down to $6.5 h^{-1}$ Mpc, makes our method very promising for the analysis of galaxy redshift surveys¹⁰. Reconstruction of the primordial positions and velocities of matter will allow us to test the gaussian nature of the primordial perturbations and the self-consistency of cosmological hypotheses, such as the choice of the global cosmological parameters and the assumed biasing scheme. By obtaining a point-by-point reconstruction of the specific realization that describes the observed patch of our Universe, we could distinguish between universal properties and the influence of the large-scale environment on the galaxy formation process. Moreover, reconstruction would open a new window not only onto the past but also into the present Universe: it would enable us to determine the peculiar velocities of a very large number of galaxies, using their positions in redshift catalogues. $\square$

Methods

Monge–Ampère equation

The lagrangian map $\mathbf{q} \mapsto \mathbf{x}$ is taken to be the gradient of a convex potential $\Phi(\mathbf{q})$; therefore its inverse $\mathbf{x} \mapsto \mathbf{q}$ also has a potential representation $\mathbf{q} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$, where $\Theta(\mathbf{x})$ is again a convex function; the two potentials are Legendre–Fenchel transforms of each other (see pages 61–65 in ref. 30):

$$\Theta(\mathbf{x}) = \max_{\mathbf{q}} [\mathbf{q} \cdot \mathbf{x} - \Phi(\mathbf{q})]; \quad \Phi(\mathbf{q}) = \max_{\mathbf{x}} [\mathbf{x} \cdot \mathbf{q} - \Theta(\mathbf{x})] \quad (3)$$

The potential Θ satisfies the Monge–Ampère equation (1), written for the first time by Ampère²⁰ by exploiting the property of the Legendre transformation. Note that within the more restricted framework of the Zel'dovich approximation, Θ differs just by a quadratic additive term from the eulerian velocity potential¹⁹.

Quadratic cost function

To show that the quadratic cost minimization leads to the Monge–Ampère equation, we define the displacement field $\xi(\mathbf{x}) \equiv \mathbf{x} - \mathbf{q}(\mathbf{x})$ and perform a variation $\delta\xi(\mathbf{x})$ to obtain, to lowest order, the variation of the cost function $\delta I = \int_{\mathbf{x}} 2\xi(\mathbf{x}) \cdot (\rho(\mathbf{x})\delta\xi) d^3x$. The condition that the eulerian density remains unchanged, which constrains the variation, is expressed as $\nabla_{\mathbf{x}} \cdot (\rho(\mathbf{x})\delta\xi(\mathbf{x})) = 0$. By a simple Lagrange multiplier argument, this implies that ξ must be a gradient of some function of $\mathbf{x}$; thus, $\mathbf{q} = \mathbf{x} - \xi = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$. Furthermore, should Θ be non-convex and thus lead to multi-streaming, this would prevent the lagrangian map from being optimal.

History of mass transportation

Monge²⁴ posed the following problem: how to optimally move material from one place to another, knowing only its initial and final spatial distributions, the cost being a prescribed function of the distance travelled by 'molecules' of material (a linear function in Monge's original work). Kantorovich²⁵ showed that Monge's query was an instance of the linear programming problem, and developed for it a theory that found numerous applications in economics and applied mathematics.

Received 26 September 2001; accepted 5 April 2002.

- Narayanan, V. K. & Croft, R. A. Recovering the primordial density fluctuations: a comparison of methods. *Astrophys. J.* **515**, 471–486 (1999).
- Peebles, P. J. E. Tracing galaxy orbits back in time. *Astrophys. J.* **344**, L53–L56 (1989).
- Weinberg, D. H. Reconstructing primordial density fluctuations—I. Method. *Mon. Not. R. Astron. Soc.* **254**, 315–342 (1992).
- Nusser, A. & Dekel, A. Tracing large-scale fluctuations back in time. *Astrophys. J.* **391**, 443–452 (1992).
- Croft, R. A. & Gaztañaga, E. Reconstruction of cosmological density and velocity fields in the Lagrangian Zel'dovich approximation. *Mon. Not. R. Astron. Soc.* **285**, 793–805 (1997).
- Nusser, A. & Branchini, E. On the least action principle in cosmology. *Mon. Not. R. Astron. Soc.* **313**, 587–595 (2000).
- Goldberg, D. M. & Spergel, D. N. Using perturbative least action to recover cosmological initial conditions. *Astrophys. J.* **544**, 21–29 (2000).
- Valentine, H., Saunders, W. & Taylor, A. Reconstructing the IRAS point source catalog redshift survey with a generalized PIZA. *Mon. Not. R. Astron. Soc.* **319**, L13–L17 (2000).
- Branchini, E., Eldar, A. & Nusser, A. Peculiar velocity reconstruction with fast action method: tests on mock redshift surveys. *Mon. Not. R. Astron. Soc.* (submitted); preprint astro-ph/0110618 at (<http://xxx.lanl.gov>) (2001).
- Frieman, J. A. & Szalay, A. S. Large-scale structure: entering the precision era. *Phys. Rep.* **333–334**, 215–232 (2000).
- Bertschinger, E. & Dekel, A. Recovering the full velocity and density fields from large-scale redshift-distance samples. *Astrophys. J.* **336**, L5–L8 (1989).
- Zel'dovich, Ya. B. Gravitational instability: an approximate theory for large density perturbations. *Astron. Astrophys.* **5**, 84–89 (1970).
- Moutarde, F., Alimi, J.-M., Bouchet, F. R., Pellat, R. & Ramani, A. Precollapse scale invariance in gravitational instability. *Astrophys. J.* **382**, 377–381 (1991).
- Buchert, T. Lagrangian theory of gravitational instability of Friedman–Lemaître cosmologies and the Zel'dovich approximation. *Mon. Not. R. Astron. Soc.* **254**, 729–737 (1992).
- Munshi, D., Sahni, V. & Starobinsky, A. Nonlinear approximations to gravitational instability: a comparison in the quasi-linear regime. *Astrophys. J.* **436**, 517–527 (1994).
- Catelan, P. Lagrangian dynamics in non-flat universes and non-linear gravitational evolution. *Mon. Not. R. Astron. Soc.* **276**, 115–124 (1995).
- Gurbatov, S. & Saichev, A. I. Probability distribution and spectra of potential hydrodynamic turbulence. *Radiophys. Quant. Electr.* **27**, 303–313 (1984).
- Shandarin, S. F. & Zel'dovich, Ya. B. The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium. *Rev. Mod. Phys.* **61**, 185–220 (1989).
- Vergassola, M., Dubrulle, B., Frisch, U. & Noullez, A. Burgers' equation, Devil's stair-cases and the mass distribution for large-scale structures. *Astron. Astrophys.* **289**, 325–356 (1994).
- Ampère, A.-M. Mémoire concernant l'Application de la Théorie exposée dans le XVII^e Cahier du Journal de l'École Polytechnique, à l'Intégration des Équations aux différentielles partielles du premier et du second ordre. *J. École. R. Polytech.* **11**, 1–188 (1820).
- Brenier, Y. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci.* **305**, 805–808 (1987).
- Gangbo, W. & McCann, R. J. The geometry of optimal transportation. *Acta Math.* **177**, 113–161 (1996).
- Benamou, J.-D. & Brenier, Y. The optimal time-continuous mass transport problem and its augmented Lagrangian numerical resolution. *Numer. Math.* **84**, 375–393 (2000) also at (<http://www.inria.fr/rrrt/rr-3356.html>).
- Monge, G. Mémoire sur la théorie des déblais et des remblais. *Hist. Acad. R. Sci. Paris*, 666–704 (1781).
- Kantorovich, L. On the translocation of masses. *C.R. Acad. Sci. URSS* **37**, 199–201 (1942).
- Rockafellar, R. T. *Convex Analysis* (Princeton Univ. Press, 1970).
- Hénon, M. A mechanical model for the transportation problem. *C.R. Acad. Sci.* **321**, 741–745 (1995).
- Balinski, M. L. A competitive (dual) simplex method for the assignment problem. *Math. Program.* **34**, 125–141 (1986).
- Couchman, H. M. P., Thomas, P. A. & Pearce, F. R. Hydra: an adaptive-mesh implementation of P³-M-SPH. *Astrophys. J.* **452**, 797–813 (1995).
- Arnold, V. I. *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).

Acknowledgements

Special thanks are due to E. Branchini (observational and conceptual aspects), Y. Brenier (mathematical aspects) and M. Hénon (algorithmic aspects and the handling of spatial periodicity and of scatter plots). We also thank J. Bec, H. Frisch, B. Gladman, L. Moscardini, A. Noullez, C. Porciani, M. Rees, E. Spiegel, A. Starobinsky and P. Thomas for comments. This work was supported by the BQR program of Observatoire de la Côte d'Azur, by the TMR program of the European Union (U.F., R.M.), by MIUR (S.M.), by the French Ministry of Education, the McDonnell Foundation, the Russian RFBR and INTAS (A.S.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.F. (e-mail: uriel@obs-nice.fr).

Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars

Michael B. Wyatt & Harry Y. McSween Jr

Department of Geological Sciences, University of Tennessee, Knoxville, Tennessee 37996, USA

Mineral abundances derived from the analysis of remotely sensed thermal emission data from Mars have been interpreted to indicate that the surface is composed of basalt (Surface Type 1) and andesite (Surface Type 2)¹. The global distribution of these rock types is divided roughly along the planetary dichotomy which separates ancient, heavily cratered crust in the southern hemisphere (basalt) from younger lowland plains in the north (andesite)¹. But the existence of such a large volume of andesite is difficult to reconcile with our present understanding of the geological evolution of Mars. Here we reinterpret martian surface rock lithologies using mineral abundances from previous work¹ and new mineralogies derived from a spectral end-member set representing minerals common in unaltered and low-temperature aqueously altered basalts. Our results continue to indicate the dominance of unaltered basalt in the southern highlands, but reveal that the northern lowlands can be interpreted as weathered basalt as an alternative to andesite. The coincidence between locations of such altered basalt and a suggested northern ocean basin implies that lowland plains material may be composed of basalts weathered under submarine conditions or weathered basaltic sediments transported into this depocentre.

The identification of basalt on Mars from Thermal Emission Spectrometer (TES) observations^{1,2} agrees with visible/near-infrared spectroscopic observations of low-albedo regions³. The identification of andesite on Mars from TES observations¹ is consistent with chemical analyses of rocks by the Mars Pathfinder⁴. Additional calibrations using a variety of unweathered terrestrial volcanic rocks have demonstrated the fidelity with which thermal emission data can distinguish basalt and andesite⁵, and these results⁶ support the identification of basaltic and andesitic martian surface compositions.

The large expanse of andesite in the northern plains as suggested by Bandfield *et al.*¹ is, however, difficult to explain. On Earth, andesite forms mostly at subduction zones, where water released from descending slabs promotes mantle melting. However, there is no evidence for subduction on Mars. Although andesite can also be produced during fractionation of basaltic magma, this process is inefficient and could not account for a large volume of andesite⁷. Less fractionation is required for hydrous basaltic melts to reach the composition of andesite, and evidence exists that some martian basaltic meteorites contained water⁸. In this case, however, the northern plains should consist of comparable amounts of andesite and basalt.

The observed geographic concentration of martian andesite within the northern lowlands¹ suggests an alternative mechanism for their formation. Previous abstracts have indicated that weathered or oxidized basaltic glass is spectrally similar to andesitic glass^{9,10}. Here, we explore whether Surface Type 2 (andesite) might be reinterpreted as basalt weathered in a submarine environment. The northern lowlands lie, for the most part, below the 0-km elevation datum, and lakes or oceans, possibly fed by large outflow channels originating in the southern highlands, may have occupied this region¹¹. Deconvolved mineral abundances from ref. 1 and new mineralogies derived from TES data and terrestrial basalts using a

spectral end-member set representing minerals common in unaltered and low-temperature aqueously altered basalts are used to reclassify martian surface lithologies.

Figure 1a and b compare spectral fits and modelled mineral abundances produced by linear deconvolutions of Surface Types 1 and 2 spectra (see Methods). Bandfield *et al.*¹ used 45 spectral end-members¹² representing igneous, sedimentary and metamorphic minerals, whereas Hamilton *et al.*⁶ used a narrower range of 29 mineral spectra representing mostly unweathered basalts and andesites. Our 39 spectral end-members represent minerals common in unaltered and low-temperature aqueously altered basalts¹³ (Table 1 of the Supplementary Information). One of the primary differences between the spectral end-member sets of refs 1 and 6 and the end-member set used here is the omission of a high-silica glass phase⁵ in

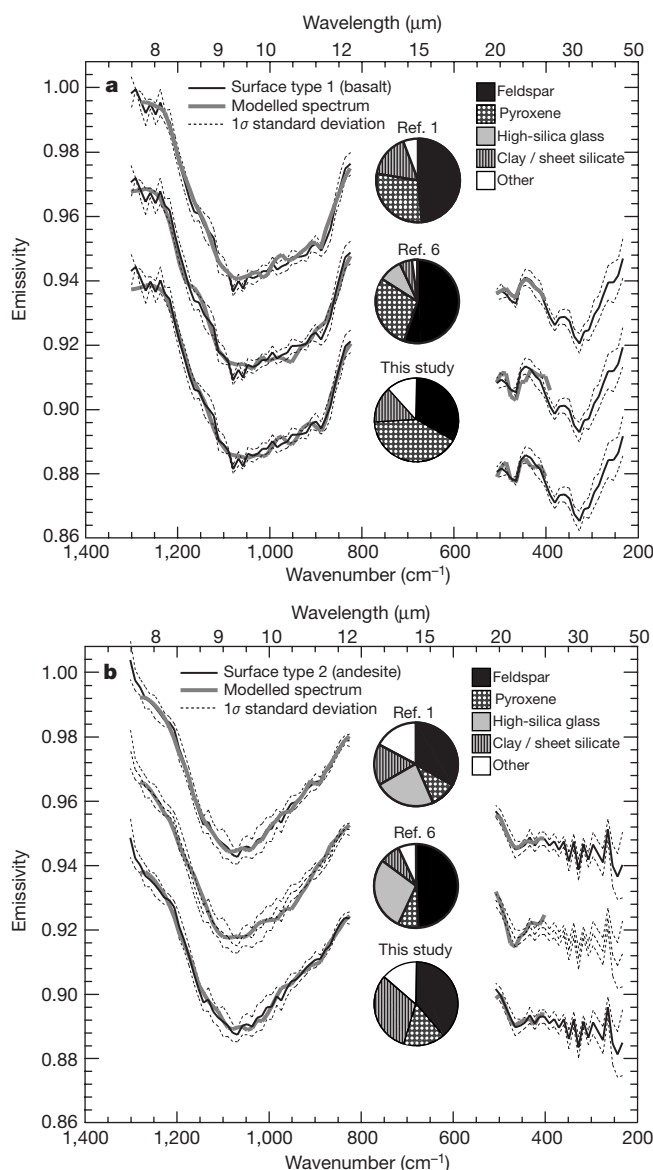


Figure 1 Comparison of martian surface spectra and modelled spectral fits. **a**, **b**, Martian Surface Type 1 (**a**) and Type 2 (**b**) spectra and modelled spectral fits (offset by 0.028 emissivity for clarity) and mineral abundances produced by linear deconvolution from Bandfield *et al.*¹, Hamilton *et al.*⁶ and this study (Table 2 of the Supplementary Information). The detection limit for mineral abundances deconvolved from TES spectra is 10–15 vol.% (ref. 2). Mineral abundances for Surface Type 2 are characteristic of either andesitic or weathered basaltic compositions.

this study (see Methods). A similar difference involving K-feldspar occurs between the end-member set used by ref. 6 compared to those used by ref. 1 and this study (see Methods). Also shown in Fig. 1a and b are 1σ standard deviations for martian surface spectra as determined by ref. 1.

Spectral fits produced by linear deconvolution of the Surface Type 1 spectrum show low root-mean-square (r.m.s.) values of 0.0018 (ref. 1), 0.0026 (ref. 6), and 0.0018 (this study). Modelled mineral abundances for the three end-member sets (Fig. 1a, Table 2 of the Supplementary Information) are all characterized by high values of plagioclase and pyroxene, and low values of weathering products (clays, sulphates and carbonates). The low r.m.s. values and similar modelled mineral abundances derived from different end-member sets indicate these modes accurately reflect the Surface Type 1 composition. The modelled mineral abundances, and the broad, slightly squared shape absorption between ~ 800 – $1,200\text{ cm}^{-1}$ characteristic of Surface Type 1, are representative of unweathered

basaltic compositions^{1,2,5,6}.

Analogous spectral fits of the Surface Type 2 spectrum show low r.m.s. values of 0.0009 (ref. 1), 0.0023 (ref. 6), and 0.0014 (this study). However, modelled mineral abundances differ, resulting in an ambiguous description of the Surface Type 2 composition (Fig. 1b, Table 2 of the Supplementary Information). The results of refs 1 and 6 are characterized by high modelled abundances of plagioclase and high-silica glass and low abundances of pyroxene and weathering products. Modelled abundances in this study are characterized by plagioclase with appreciable alteration minerals (clays and K-feldspar) and lesser pyroxene. The well-modelled spectra indicate that each of the different mineral modes is equally valid for describing the Surface Type 2 spectrum. Modelled mineral abundances from refs 1 and 6, and the slightly V-shaped absorption between ~ 800 – $1,200\text{ cm}^{-1}$ characteristic of Surface Type 2, are representative of andesite^{1,5,6}. Higher abundances of clays and K-feldspar, and lower abundances of pyroxenes for Surface Type 2 as modelled in this study, are characteristic of altered basalt¹³.

Modelled mineral abundances supporting a weathered basalt composition are largely a result of clays replacing high-silica glass. Absorption features between 500 – 550 cm^{-1} in laboratory spectra distinguish clays modelled in this study (Fe-smectite and Ca-montmorillonite) from high-silica glass (Fig. 2a); however, the CO_2 atmosphere of Mars is largely opaque in this spectral region, which limits its usefulness for analysis of surface compositions. The spectral regions of high-silica glass and clays used for analysis of martian surface compositions are very similar in their overall spectral shape and positions of spectral features. Future work is needed to understand better the degree of substitution and spectral effects of varying crystallinity in amorphous phases over a range of chemical compositions. Despite this ambiguity, conclusions from deconvolved TES spectra concerning other rock-forming minerals are thought to be robust^{1,2,5,6}.

Figure 2b shows spectra of the fresh-cut and weathered surfaces of a Columbia River Flood Basalt (CRB sample 4, refs 5, 6) re-sampled to TES spectral resolution compared to Surface Type 1 and 2 spectra, respectively. Because martian surface spectra analysed by TES are thought to be dominated by sand-sized particles^{12,1} rather than bedrock, a blackbody component (which does not affect spectral shape) was added to each terrestrial sample to account for band-depth particle-size effects. The spectrum of the fresh-cut CRB surface displays a broad, slightly squared shape absorption between ~ 800 – $1,200\text{ cm}^{-1}$ similar to the Surface Type 1 spectrum, whereas the spectrum of the natural CRB surface displays a more V-shaped absorption similar to the Surface Type 2 spectrum. Spectral fits produced by linear deconvolution of the fresh-cut CRB surface spectrum show low r.m.s. values (0.0020 (ref. 1), 0.0025 (ref. 6), 0.0031 (this study)), and deconvolved mineral abundances are characterized by high values of plagioclase and pyroxene, and low values of weathering products (Table 3 of the Supplementary Information), similar to Surface Type 1. Spectral fits produced by linear deconvolution of the natural CRB surface spectrum show analogous low r.m.s. values (0.0023 (ref. 1), 0.0034 (ref. 6), 0.0030 (this study)). However, modelled mineral abundances for the natural CRB surface using the end-member sets of Bandfield *et al.*¹ and Hamilton *et al.*⁶ are characterized by high abundances of plagioclase, pyroxene and high-silica glass, whereas abundances in this study are characterized by plagioclase with appreciable alteration minerals and lesser pyroxene (Table 3 of the Supplementary Information), similar to the ambiguous mineralogical characterization of Surface Type 2.

Microscopic examination of the fresh-cut and natural CRB surfaces agrees best with deconvolved mineral abundances from Bandfield *et al.*¹ in which a combination of plagioclase, pyroxene, high-silica glass and clays comprises the sample's modal mineralogy. A decrease in pyroxene and increase in clay abundances is expected on a natural surface as pyroxenes are typically weathered to clays.

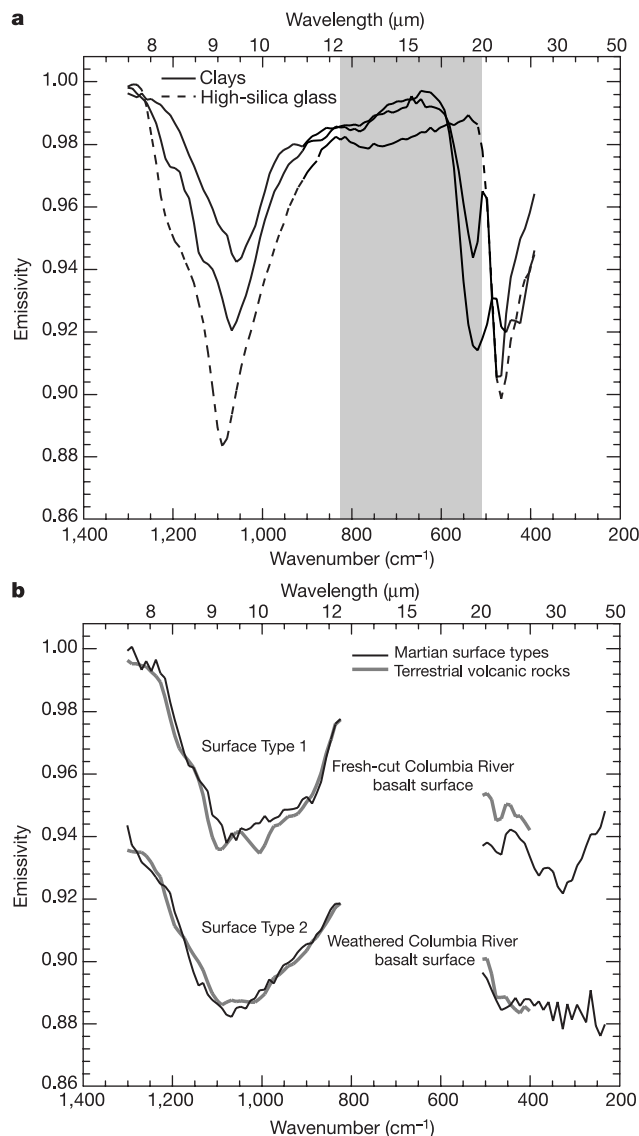


Figure 2 Spectral comparisons of glass, clays, terrestrial basalt and martian surface spectra. **a**, **b**, High-silica glass and two clays (Fe-smectite and Ca-montmorillonite) (**a**) and fresh-cut and weathered surface of a CRB sample^{5,6} (**b**) compared to Surface Type 1 and 2 spectra, respectively (offset by 0.06 emissivity for clarity). Blackbody was added to each terrestrial rock sample to account for band-depth particle-size effects.

However, an increase in the abundance of high-silica volcanic glass on a natural surface is not expected, because such a phase is also easily weathered. The increase in modelled high-silica glass from the fresh-cut to the natural CRB surface might indicate the formation of amorphous silica alteration products which are spectrally similar to high-silica glass. The original interpretation of the high-silica glass spectral end-member phase as a primary volcanic glass may be too limited, as it could also represent an amorphous high-silica alteration product.

Figure 3 compares the global distribution of Surface Types 1 and 2, as determined in ref. 1 and global topography determined from Mars Orbiter Laser Altimeter (MOLA) data¹⁴ in hemispherical north and south stereographic polar projections. The highest concentrations of Surface Type 1 (green) in the southern highlands and Surface Type 2 (red) in the northern lowlands coincide roughly with the planetary dichotomy, although there is some mixing of surface types across this divide. Blue pixels on TES composition maps represent regions covered by fine-grained bright dust which blankets the surface and prohibits spectral analysis of sand and rock compositions. Also shown on the north polar projections is a white line defining contacts previously interpreted as shorelines¹⁵ which approximate an equipotential line¹¹. Surface Type 2 compositions in the northern hemisphere are distributed almost entirely within the proposed shorelines where they are not spectrally obscured by dust cover. Surface Type 2 compositions in the southern hemisphere may be due to similar alteration processes, but the lack of geographic

concentrations does not support a large southern hemisphere body of water.

Linear deconvolution of the Surface Type 2 spectrum by ref. 1 and this study requires some clay/sheet silicates to be modelled at or above the TES detectability limit². The high abundance of modelled alteration products in this study results from the spectral similarities and substitution of clays for high-silica glass at TES spectral resolution. High abundances of well crystalline phyllosilicates in low-albedo regions are not supported by Phobos Infrared Spectrometer (ISM) data¹⁶; however, poorly crystalline palagonite has been proposed as a component in bright martian soil¹⁷. We note that ISM data only covered ~20% of the martian equatorial region, which is dominated by unweathered basaltic compositions as determined by TES data. Comparison of ISM data with weathered basalt compositions in the northern hemisphere is thus not possible. Merényi *et al.*¹⁸ did observe northern regions from high-spectral-resolution telescopic images and noted that weaker mafic mineral bands could be explained by a higher glass content. This is generally analogous to the increased glass content determined¹ (interpreted here as a possible alteration product) to explain differences in the Surface Type 1 and 2 spectra. The mineralogies determined by Bandfield *et al.*¹ and this study for Surface Type 2 can thus both be interpreted as representing weathered basalt.

A martian crust dominated by basalt containing augite and pigeonite (Table 3 of the Supplementary Information) is in agreement with visible/near-infrared spectroscopic observations of low-

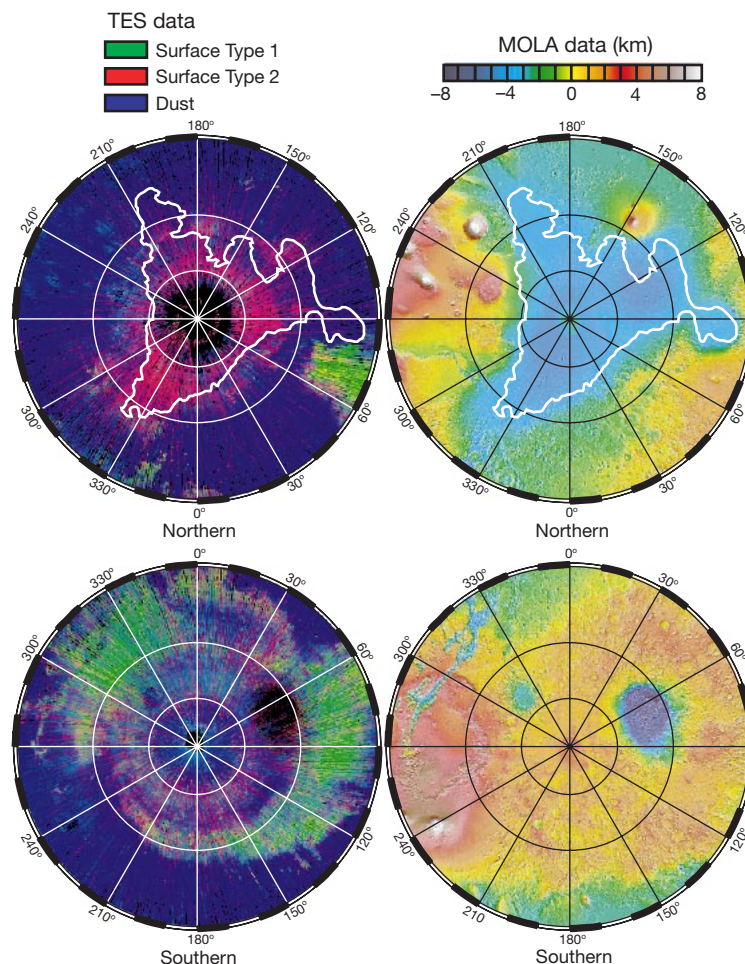


Figure 3 Global distribution of Surface Types 1 and 2 as determined by ref. 1 and global topography determined from MOLA data¹⁴. Surface Type 2 compositions in the northern

hemisphere map closely to a proposed ocean basin¹¹, as delineated by elevation and a possible shoreline (white line)¹⁵.

albedo regions that suggest surface materials may be similar to shergottite meteorites³. Analysis of depressions (“ghost craters”) in the lowlands indicate that the basement there is of Noachian age¹⁹, like the southern highlands, so it is plausible that the entire planet had an ancient basaltic crust that was altered where water ponded. This hypothesis may also explain the absence of andesite among martian meteorites, because either this material might exist only as weathering rinds or more thoroughly weathered rocks might not survive impact ejection. McSween and Keil²⁰ compared compositions of martian soils and concluded that the global dust formed by weathering of basalts rather than andesites. Although the existence of andesite on Mars remains a viable hypothesis, linear deconvolution of martian thermal infrared spectra, spectral matching with terrestrial volcanic rocks, and the geographic distribution of the Surface Type 2 spectrum suggest that martian northern lowland plains materials are basalts weathered under submarine conditions and/or sediments derived from weathered basalt and deposited in this basin. □

Methods

Linear deconvolutions using three spectral end-member sets¹² (refs 1, 6 and this study) were run for each martian surface spectrum. Algorithm outputs include a best-fit model spectrum for each martian surface spectrum, percentage of each mineral end-member used, and an r.m.s. error value (average error over the entire spectrum). The r.m.s. value indicates goodness of fit for a particular model iteration for the same martian spectrum, rather than a comparison between fits of the two martian spectra. Spectral fitting for both methods was constrained to 1,280–400 cm⁻¹, although TES data cover the wavenumber range of 1650–200 cm⁻¹. The atmospheric correction used to derive the TES surface spectra used in this study did not include the high wavenumber range of TES data owing to numerous water vapour and minor CO₂ features. Furthermore, there are no fundamental silicate features in the 1,650–1,400 cm⁻¹ region. Similar to ref. 1 the deconvolution algorithm also did not use the 400–200 cm⁻¹ spectral range. Residual atmospheric water vapour rotational bands are apparent in both the martian surface spectra as well as in the Arizona State University (ASU) mineral library end-member spectra. This restricted wavenumber range prevents the deconvolution algorithm from attempting to fit water vapour features instead of surface mineralogy.

Accurate deconvolution is largely dependent on spectral end-member sets representing an appropriate range of compositions in the mixture. The high-silica glass phase (77.9% SiO₂, 12.6% Al₂O₃, 3.90% Na₂O, 5.67% K₂O) not included in the end-member set used in this study has been interpreted^{15,6} as representing a primary volcanic glass. High-silica volcanic glasses are often not present in high abundances in unweathered basalts, compared to andesites, and typically alter to clays in a weathering environment. For these reasons, we have excluded this volcanic glass phase in our end-member set that is designed to represent basalts and low-temperature aqueously altered basalts. K-feldspar is not included in the Hamilton *et al.*⁶ spectral end-member set as it is not typically found in unweathered basalt and andesite. K-feldspar is however included by Bandfield *et al.*¹ to represent possible mineralogies in a larger suite of igneous, metamorphic and sedimentary rock types and in this study as a possible alteration product¹³.

Received 26 September 2001; accepted 15 March 2002.

- Bandfield, J. L., Hamilton, V. E. & Christensen, P. R. A global view of martian surface compositions from MGS-TES. *Science* **287**, 1626–1630 (2000).
- Christensen, P. R. *et al.* Identification of a basaltic component on the martian surface from Thermal Emission Spectrometer data. *J. Geophys. Res.* **105**, 9609–9621 (2000).
- Mustard, J. F., Murchie, S., Erard, S. & Sunshine, J. M. In situ compositions of martian volcanics: Implications for the mantle. *J. Geophys. Res.* **102**, 25605–25615 (1997).
- Rieder, R. *et al.* The chemical composition of martian soils and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* **278**, 1771–1774 (1997).
- Wyatt, M. B. *et al.* Analysis of terrestrial and martian volcanic compositions using thermal emission spectroscopy: I. Determination of mineralogy, chemistry, and classification strategies. *J. Geophys. Res.* **106**, 14711–14732 (2001).
- Hamilton, V. E. *et al.* Analysis of terrestrial and martian volcanic compositions using thermal emission spectroscopy: II. Application to martian surface spectra from the Mars Global Surveyor Thermal Emission Spectrometer. *J. Geophys. Res.* **106**, 14733–14746 (2001).
- McSween, H. Y. Jr *et al.* Chemical, multispectral, and textural constraints on the composition and origin of rocks at the Mars Pathfinder landing site. *J. Geophys. Res.* **104**, 8679–8715 (1999).
- McSween, H. Y. Jr *et al.* Geochemical evidence for magmatic water within Mars from pyroxenes in the Shergotty meteorite. *Nature* **409**, 487–490 (2001).
- Noble, S. K. & Pieters, C. M. Type 2 terrain: compositional constraints on the martian lowlands. *Lunar Planet. Sci. Conf. XXXII*, [online] CD 1230 (2001).
- Minitti, M. E., Rutherford, M. J. & Mustard, J. F. The effects of oxidation on spectra of SNC-like basalts: Application to Mars remote sensing. *Lunar Planet. Sci. Conf. XXXI*, [online] CD 1282 (2000).
- Head, J. W. *et al.* Possible ancient oceans on Mars; evidence from Mars Orbiter laser altimeter data. *Science* **286**, 2134–2137 (1999).
- Christensen, P. R. *et al.* A thermal emission spectral library of rock forming minerals. *J. Geophys. Res.* **105**, 9735–9739 (2000).
- Thompson, G. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Bostrom, K., Laubier, L. & Smith, K. L. Jr) 225–278 (Plenum, New York, 1983).

- Smith, D. E. *et al.* The global topography of Mars and implications for surface evolution. *Science* **284**, 1495–1503 (1999).
- Parker, T. J. *et al.* Transitional morphology in the west Deuteronilus Mensae region of Mars: implications for modification of the lowland/upland boundary. *Icarus* **82**, 111–145 (1989).
- Murchie, S., Kirkland, L., Erard, S., Mustard, J. & Robinson, M. Near-infrared spectral variations of martian surface materials from ISM imaging spectrometer data. *Icarus* **147**, 444–471 (2000).
- Murchie, S. *et al.* Spatial variations in the spectral properties of bright regions on Mars. *Icarus* **105**, 454–468 (1993).
- Merényi, E., Singer, R. B. & Miller, J. S. Mapping of spectral variations on the surface of Mars from high spectral resolution telescopic images. *Icarus* **124**, 280–295 (1996).
- Frey, H. V. *et al.* A very large population of likely buried impact basins in the northern lowlands of Mars revealed by MOLA data. *Lunar Planet. Sci. Conf. XXXII*, [online] CD 1680 (2001).
- McSween, H. Y. Jr & Keil, K. Mixing relationships in the martian regolith and the composition of globally homogenous dust. *Geochim. Cosmochim. Acta* **64**, 2155–2166 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Bandfield, J. Moersch, T. Hare, J. Mustard, R. Clark, V. Hamilton and P. Christensen for discussions of the results presented here.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.Y.M. (e-mail: mcsween@utk.edu).

Formation of tough interlocking microstructures in silicon nitride ceramics by dynamic ripening

Zhijian Shen, Zhe Zhao, Hong Peng & Mats Nygren

Department of Inorganic Chemistry, BRIIE Center for Inorganic Interfacial Engineering, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

Ceramics based on Si₃N₄ have been comprehensively studied and are widely used in structural applications^{1,2}. The development of an interlocking microstructure of elongated grains is vital to ensure that this family of ceramics have good damage tolerance^{3,4}. Until now this has been accomplished by heating the appropriate powder compacts to temperatures above 1,700 °C for extended periods. This procedure involves a necessary step of controlling the size and population of seeds—added *ex situ* or formed *in situ*—to ensure selective grain growth^{5,6}. Here we report the very fast (within minutes) *in situ* formation of a tough interlocking microstructure in Si₃N₄-based ceramics. The microstructures are obtained by a dynamic ripening mechanism, an anisotropic Ostwald ripening process that results from the rapid heating rate. The resulting microstructures are uniform and reproducible in terms of grain size distribution and mechanical properties, and are easily tailored by manipulating the kinetics. This process is very efficient and opens up new possibilities to optimize mechanical properties and cost-effectively manufacture ceramics.

Sintering Si₃N₄-based ceramics is a complex process, during which densification, phase transformation and grain growth take place simultaneously. Therefore, it is not an easy task to clarify the mechanisms determining the rate of grain growth, for example. Research carried out during recent years indicates that the grain growth is dominated by an Ostwald ripening mechanism driven by the reduction of interface energy, that is, the smallest grains deriving from the precursor powder and early precipitates dissolve in the

liquid formed at high temperatures, and the appropriate constituents (ionic species and/or clusters) are transported in the liquid to larger grains where they re-precipitate. Efforts have thus been made to use this mechanism to form a tough interlocking microstructure of elongated grains by means of: (1) seeding; (2) varying the amount and composition of co-existing liquid; (3) stimulating the dissolution process by using fine-grained precursor powders; (4) controlling nucleation through phase transformation^{5,7,8}. Although it has been observed that under high nitrogen pressure a few grains grow extremely fast above 1,900 °C (refs 9, 10), and that the growth of anisotropic grains is directly related to the difference in solubility between small and large grains¹¹, it is generally accepted that growth of elongated grains is predominantly diffusion-controlled¹². This implies that extended holding at high temperature is a necessary requirement to develop a tough interlocking microstructure in any process that is based on such a principle of selective grain growth.

We have used a rapid consolidation technique, spark plasma sintering (SPS), to densify a variety of ceramic and powdered-metal materials. For Si₃N₄-based ceramics we have observed the densification and appropriate phase transformation during this process to occur within a narrow temperature interval and to be completed very rapidly, typically between 1,350 and 1,550 °C and within 1 to 2 min. The use of fast heating rates implies that: (1) formation of

unwanted crystalline intermediate phases can be avoided; (2) the grain growth in the low temperature region is suppressed; (3) the locally formed liquid is rapidly homogenized via the large capillary forces present between submicrometre and/or nano-sized grains. We obtained fully-dense compacts consisting of submicrometre-sized, nearly equiaxed, grains; this provided us with the opportunity to follow the grain growth taking place during subsequent heating and annealing at temperatures ranging from 1,600 to 1,800 °C.

The width and length of randomly selected equiaxed and elongated (when present) grains in a sample of β -sialon, a solid solution based on β -Si₃N₄, heated to temperatures ranging from 1,500 to 1,800 °C are given in Fig. 1a. These samples, which have an overall composition of Y_{0.096}Si_{1.663}Al_{0.222}O_{0.509}N_{2.196} and contain approximately 20 vol.% of an extra liquid/glassy phase, were heated at a rate of 200 °C min⁻¹ with no holding at final sintering temperature under an uniaxial pressure of 50 MPa. Corresponding data for isothermal grain growth at 1,600 °C are given in Fig. 1b. The insets show that the fracture toughness, as expected, increases with the formation of elongated grains. The microstructures of the samples sintered at 1,600 °C without and with 1 and 10 min holding times are shown in Fig. 2. It is very surprising that the anisotropic grain growth progresses so rapidly that the nearly equiaxed submicrometre-sized grains grow to well-developed elongated grains within 1 min. A tough interlocking microstructure was developed and every grain shows a clear tendency to develop elongated morphology. The grain growth process appears to be thermally activated, and for this particular composition the observed critical temperature is ~1,600 °C, because the samples prepared at 1,550 °C did not contain any elongated grains even after prolonged heat treatment at this temperature.

Similar fast grain growth has been observed in compacts of α -sialon, a solid solution based on α -Si₃N₄, both with stoichiometric compositions and with extra liquid phase added. For example, using α -Si₃N₄ as precursor powder, the stoichiometric ytterbium-stabilized α -sialon Yb_{0.35}Si_{9.75}Al_{2.25}O_{1.2}N_{14.8} was densified, without holding, at 1,700 and 1,750 °C under a pressure of 50 MPa by applying a heating rate of 200 °C min⁻¹. X-ray diffraction and SEM structural analyses revealed complete phase transformation in both cases, with the former preparation consisting of equiaxed submicrometre grains and the latter of elongated grains up to 10 μ m in length. The two materials showed similar hardness (22 GPa), but the latter is tougher than the former, with fracture toughness values of 3.5 and 5.5 MPa m^{1/2}, respectively. Using the same composition and otherwise identical experimental conditions but with a heating rate of 40 °C min⁻¹ yielded, at 1,750 °C, compacts containing equiaxed grains; see Fig. 3.

It is well known that the elongated growth takes place along the *c* direction of the unit cell of the α - and β -phases. It is commonly assumed that this is due to anisotropic interfacial energies. We ascribe the rapid grain growth described above to an anisotropic Ostwald ripening process that takes place dynamically—abbreviated as ‘dynamic ripening’. The difference between dynamic and steady-state Ostwald ripening is as follows. In the former case, owing to the fast heating rate applied, the chemical composition of the co-existing liquid phase deviates from that set by thermodynamic equilibrium. During the short time span when the liquid approaches equilibrium state, the dissolution of small grains is promoted so as to produce a momentarily supersaturated liquid, that is, the liquid is supplied with the appropriate constituents in a very fast and efficient way. The supersaturated liquid immediately responds by preferentially depositing these constituents onto the crystallographic surfaces that will most easily accommodate them. After this first stage of very fast dissolution of small grains the growth rate will decrease and finally the steady-state Ostwald ripening mechanism will govern the growth. The evolution of the microstructures shown in Fig. 2 with time is thus interpreted in

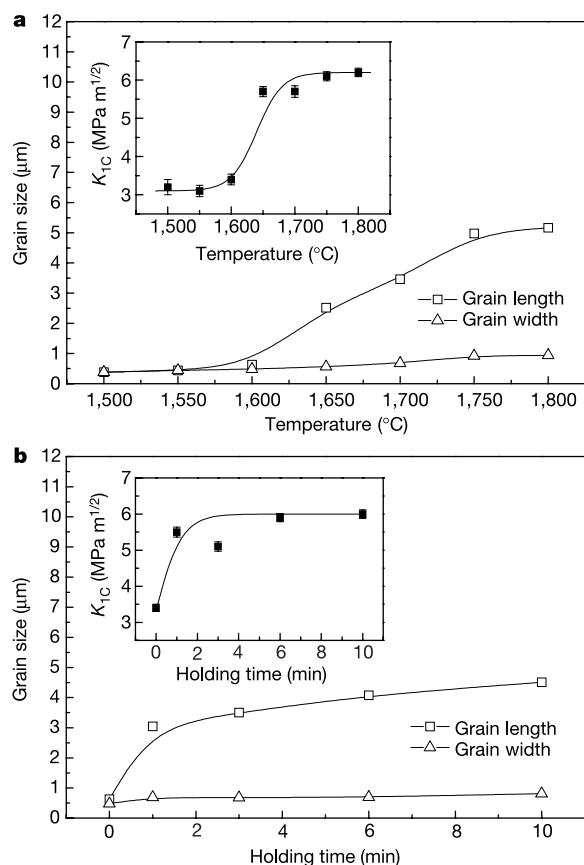


Figure 1 Width and length of randomly selected equiaxed and elongated (when present) β -sialon grains. **a**, Samples heated to temperatures ranging from 1,500 to 1,800 °C using a heating rate of 200 °C min⁻¹. **b**, Samples isothermally heated at 1,600 °C. The insets show the variation of the fracture toughness (K_{IC}) with the sintering temperature and holding time, respectively. We note the extremely fast formation of elongated grains and the evolution of the fracture toughness with the formation of elongated grains.

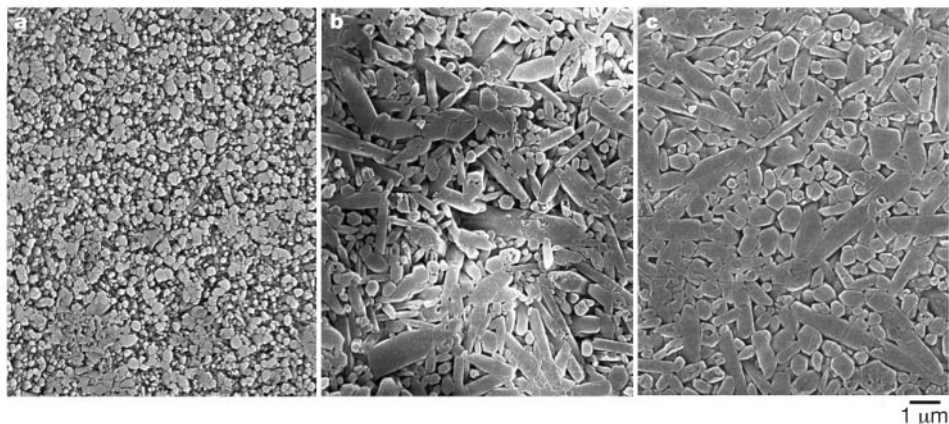


Figure 2 Scanning electron microscopy pictures of fully dense β -sialon ceramic sintered at 1,600 °C. **a**, Sample sintered without any holding time at 1,600 °C. **b**, Sample sintered with a holding of 1 min at 1,600 °C. **c**, Sample sintered with a holding of 10 min at 1,600 °C. It can also be noted that the grain size distribution formed in our experiments is

much more narrow than that obtained by the techniques based on the concept of selective grain growth. A more detailed description of the evolution of the microstructure with the holding time and sintering temperature is given in the Supplementary Information.

terms of the conditions allowing the dynamic ripening process to take place, that is, when the liquid and solids present are thermodynamically far from their equilibrium state, dynamic ripening occurs at the very beginning of the annealing time, whereas a prolonged annealing time implies that the process is transformed from dynamic ripening to steady-state ripening.

α -Sialon ceramics with similar composition have been prepared previously, following the traditional method of selective grain growth. It was concluded that it is difficult to obtain a tough interlocking microstructure of elongated grains without seeding by using α - Si_3N_4 as precursor^{5,6}. As shown above, however, anisotropic grain growth, and hence also tough interlocking microstructures of α -sialon ceramics, can be achieved by using fast heating rates without any seeding.

The use of fast heating rates is essential for obtaining the required elongated grains because slow heating ($40^\circ\text{C min}^{-1}$), a heating rate typically used in conventional hot-pressing of Si_3N_4 -based ceramics, did not yield any elongated grains, that is, the dynamic

dissolution and re-precipitation state described above is never established. We have also succeeded so far in producing such tough interlocking microstructures in β - Si_3N_4 and O-sialon ceramics, using a processing concept similar to that described above, which suggests that improved microstructures also can be achieved in other ceramic systems that are produced via a liquid-phase sintering mechanism. □

Methods

Preparation and characterization

See Supplementary Information.

Compacting

The samples were sintered in vacuum in an SPS apparatus, Dr. Sinter 2050 (Sumitomo Coal Mining). The powder precursors were loaded in cylindrical carbon dies with an inner diameter of 20 mm. The samples were heated via a pulsed d.c. current that passes through the pressure die, that is, the pressure die also acts as a heating source and this arrangement allows the use of very fast heating rates (up to $600^\circ\text{C min}^{-1}$). The temperature was automatically raised to 600 °C over a period of 3 min, and from this point and onwards it was monitored and regulated by an optical pyrometer focused on the surface of the die. A heating rate of $200^\circ\text{C min}^{-1}$ was used if not otherwise stated, and a uniaxial pressure of 50 MPa was applied from the start to the end of the sintering cycle. No holding time was applied, if not otherwise stated. The set-up allows a cooling rate of about $400^\circ\text{C min}^{-1}$ in the temperature range 1,800–1,000 °C.

An inherent property of the SPS process is that the sample is exposed to a higher temperature than the recorded one. The size of this temperature difference depends strongly on the shape of the die and also on the density, electrical and thermal conductivity of the die and sample, respectively, the pressure applied, the effectiveness of used heating shields, and so on. The temperature difference between the inner part of the sample and the surface of the die has in the present case been estimated from the melting behaviour of pieces of Pt-wires embedded in Si_3N_4 and BN powder compacts that were heated to temperatures ranging from 1,500 to 1,750 °C with no holding time at the final temperature and using a heating rate of 200 and $40^\circ\text{C min}^{-1}$, respectively. These experiments clearly suggest that in the present case the temperature difference between the surface of the die (recorded temperature) and the interior of the sample is not larger than 75 °C and that this difference is of the same order independently of the heating rate used. Further details concerning these experiments are in the Supplementary Information.

Determination of grain size

Sintered compacts of the β -sialon samples were polished down to a 1- μm finish and chemically etched in a molten mixture of KOH and KCl to expose grain boundaries. SEM pictures with fixed magnification of five randomly selected areas of each sample were recorded. From each picture ten grains with their elongated direction parallel or almost parallel to the polished surface were randomly selected and their width and length were measured. The average values of the widths and lengths of the measured grains are presented in Fig. 1.

Received 7 January; accepted 19 March 2002.

1. Jack, K. H. in *Proc. 6th Int. Symp. on Ceramic Materials and Components for Engines* (ed. Niihara, K.) 203–207 (Japan Fine Ceramics Association, Tokyo, 1997).
2. Thompson, D. P. Tough cookery. *Nature* **389**, 675–677 (1997).

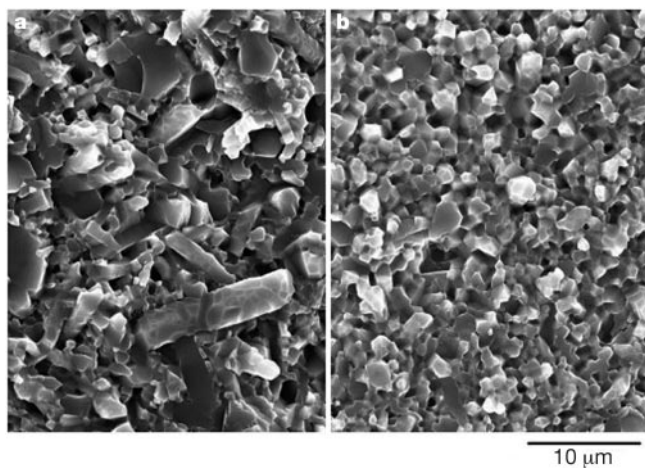


Figure 3 Scanning electron microscopy pictures of fractured surfaces of an α -sialon ceramic with a stoichiometric composition of $\text{Yb}_{0.35}\text{Si}_{9.75}\text{Al}_{2.25}\text{O}_{1.2}\text{N}_{14.8}$ heated to 1,750 °C. **a**, Sample heated using a heating rate of $200^\circ\text{C min}^{-1}$. **b**, Sample heated using a heating rate of $40^\circ\text{C min}^{-1}$. We note that the use of a high heating rate gives rise to the formation of elongated grains, whereas a slow heating rate yields equiaxed grains.

- Lange, F. F. Relations between strength, fracture energy, and microstructure of hot pressed Si_3N_4 . *J. Am. Ceram. Soc.* **56**, 518–522 (1973).
- Tani, E., Umebayashi, S., Kishi, K. & Kobayashi, K. Gas-pressure sintering of Si_3N_4 with concurrent addition of Al_2O_3 and 5wt% rare-earth oxide: high fracture toughness Si_3N_4 with fiber-like structure. *Am. Ceram. Soc. Bull.* **65**, 1311–1315 (1986).
- Chen, L.-W. & Rosenflanz, A. A tough SiAlON ceramic based on $\alpha\text{-Si}_3\text{N}_4$ with a whisker-like microstructure. *Nature* **389**, 701–704 (1997).
- Kim, J., Rosenflanz, A. & Chen, L.-W. Microstructure control of in-situ-toughened $\alpha\text{-SiAlON}$ ceramics. *J. Am. Ceram. Soc.* **83**, 1819–1821 (2000).
- Hirao, K., Nagaoka, T., Brito, M. E. & Kanzaki, S. Microstructural control of silicon nitride by seeding with rodlike β -silicon nitride particles. *J. Am. Ceram. Soc.* **77**, 1857–1862 (1994).
- Mitomo, M. in *Proc. 6th Int. Symp. on Ceramic Materials and Components for Engines* (ed. Niihara, K.) 85–90 (Japan Fine Ceramics Association, Tokyo, 1997).
- Park, D. S., Lee, S.-Y., Kim, H.-D., Yoo, B.-J. & Kim, B.-A. Extra large grains in the silicon nitride ceramics doped with yttria and hafnia. *J. Am. Ceram. Soc.* **81**, 1876–1880 (1998).
- Zhang, C., Komeya, K., Tatami, J. & Meguro, T. Inhomogeneous grain growth and elongation of Dy- α sialon ceramics at temperature above 1800 °C. *J. Eur. Ceram. Soc.* **20**, 939–944 (2000).
- Emoto, H. & Mitomo, M. Control and characterization of abnormally grown grains in silicon nitride ceramics. *J. Eur. Ceram. Soc.* **17**, 797–804 (1997).
- Kitayama, M., Hirao, K., Toriyama, M. & Kanzaki, S. Modeling and simulation of grain growth in Si_3N_4 —anisotropic Ostwald ripening. *Acta Mater.* **46**, 6541–6550 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z. S. (e-mail: shen@inorg.su.se).

Discrete stages in the solvation and ionization of hydrogen chloride adsorbed on ice particles

J. P. Devlin*, N. Uras*, J. Sadlej† & V. Buch‡

* Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74078, USA

† Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw; and Drug Institute, 00-725 Warsaw, Poland

‡ Fritz Haber Institute for Molecular Dynamics, Hebrew University, Jerusalem, Israel 91904

Ionization and dissociation reactions play a fundamental role in aqueous chemistry. A basic and well-understood example is the reaction between hydrogen chloride (HCl) and water to form chloride ions (Cl^-) and hydrated protons (H_3O^+ or H_5O_2^+). This acid ionization process also occurs in small water clusters^{1–4} and on ice surfaces^{5–17}, and recent attention has focused on the mechanism of this reaction in confined-water media and the extent of solvation needed for it to proceed^{1–4,9,15–17}. In fact, the transformation of HCl adsorbed on ice surfaces from a predominantly molecular form to ionic species during heating from 50 to 140 K has been observed^{8,13,14}. But the molecular details of this process remain poorly understood. Here we report infrared transmission spectroscopic signatures of distinct stages in the solvation and ionization of HCl adsorbed on ice nanoparticles kept at progressively higher temperatures. By using Monte Carlo and *ab initio* simulations to interpret the spectra, we are able to identify slightly stretched HCl molecules, strongly stretched molecules on the verge of ionization, contact ion pairs comprising H_3O^+ and Cl^- , and an ionic surface phase rich in Zundel ions, H_5O_2^+ .

Our experiments focused on the Fourier-transform infrared (FTIR) transmission spectroscopy of HCl adsorbate on ice nanoparticles. Three-dimensional arrays of ice nanocrystals of a 12-nm

average diameter were deposited on the windows of the infrared cold-cluster cell¹⁸. Hydrogen chloride adsorbate was introduced by alternating the loading of ice particles with the loading of much smaller (~ 1 nm) clusters of HCl or DCl. These acid clusters vaporize in the range 50–65 K, giving HCl adsorbed on the surfaces of the ice nanocrystals within the array (with $\leq 30\%$ surface coverage). At least 4 samples each were prepared of HCl adsorbed on H_2O ice, DCl adsorbed on D_2O ice, and HCl adsorbed on D_2O ice. Consistent results were also obtained for HBr on H_2O . The main experimental results are the difference spectra obtained for acid-covered nanocrystals and bare nanocrystals prepared under similar conditions. This difference method eliminates the absorption bands associated with ice in the interior of the nanocrystal, and thus leaves the weaker bands due to adsorbed acids clearly visible.

The difference spectra of HCl adsorbed on H_2O and D_2O nanoparticles at 50 K reveal two distinct HCl bands with bandwidths of ~ 150 cm^{-1} centred at 2,500 and 1,710 cm^{-1} (Fig. 1). The corresponding frequency shifts with respect to gaseous HCl are -385 and $-1,175$ cm^{-1} , suggesting that the bands correspond respectively to predominantly covalent, hydrogen-bonded molecules, and to strongly stretched molecules on the verge of ionization. Figure 2 shows analogous difference spectra for a fully deuterated system at 60, 80 and 90 K, with the changes in the spectrum with increasing temperature attributed to progressive ionization (see below). The spectra also exhibit a continuum absorption, which grows with temperature (Fig. 2, top panel)—this is a signature of the Zundel ion¹⁹.

To identify the surface adsorption sites available to HCl on ice and assign the spectra, a 50 K Monte Carlo simulation was first carried out for molecular HCl adsorbed on an ice particle, yielding molecular configurations adopted by the adsorbate on

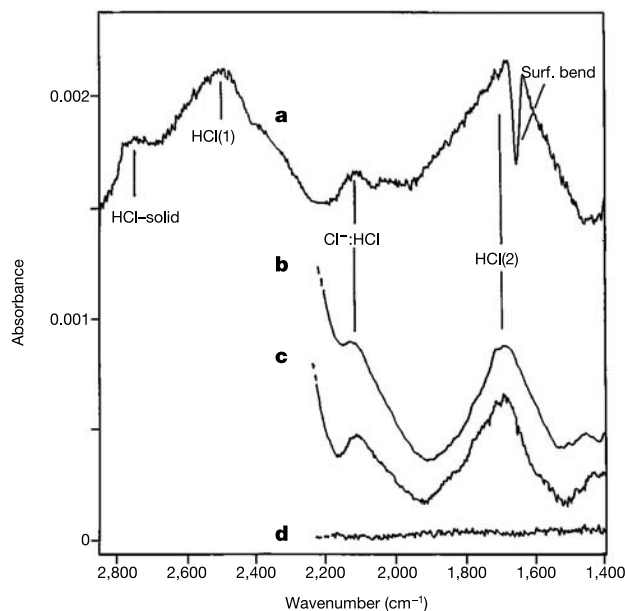


Figure 1 FTIR difference spectra of HCl on ice particles at 50 K and $\sim 15\%$ coverage. The top spectrum (trace **a**) is for HCl on H_2O ice particles, and exhibits two molecular adsorbate bands, HCl(1) and HCl(2), and the residual acid nanoparticle band, HCl-solid. Traces **b** and **c** are spectra for HCl on D_2O ice and give a clearer view of the HCl(1) band because the interfering bending motion of surface water is downshifted, away from the HCl(2) band. The use of D_2O ice also eliminates possible contamination of this band by protonated-water absorption. Traces **b** and **c** are spectra of the same HCl-ice array but referenced to different bare D_2O ice samples; trace **b** has been smoothed. Trace **d**, a difference spectrum comparing two pure D_2O ice samples, illustrates the instrumental noise.

the ice-particle surface. These configurations then served as a guide to construct smaller cluster models of the form $(\text{HCl})_m(\text{H}_2\text{O})_n$, $m = 1-2$; $n = 5-7$ (Fig. 3), which were amenable to *ab initio* calculations of a sufficiently high level to estimate the extent of HCl stretching by solvation, and the corresponding frequency shifts.

The $\text{H}_2\text{O} \cdots \text{H}_2\text{O}$ and $\text{HCl} \cdots \text{HCl}$ interactions used in the Monte Carlo simulations were adopted from refs 20 and 21, respectively; the $\text{H}_2\text{O} \cdots \text{HCl}$ potential was constructed as a sum of coulombic and Lennard-Jones terms. The ice particle model $(\text{H}_2\text{O})_{293}$ was adopted from ref. 22. Our past studies^{22,23} showed that ice nanoparticles are characterized by a crystalline interior and a disordered surface, the structure of which is dictated by the need to terminate the crystalline interior in an approximately spherical geometry. The flexibility of the hydrogen bonds is sufficient for most surface H_2O molecules to acquire four (often distorted) hydrogen bonds. Monte Carlo simulations were carried out for coverages of 18 and 79 HCl.

Molecular HCl is a good proton donor and a poor proton acceptor that bonds preferentially to dangling-O sites on the ice surface, that is, the oxygen atoms of water molecules with less than two acceptor hydrogen bonds. At the lower coverage, two-thirds of the adsorbate molecules are singly coordinated to dangling-O (Fig. 3a), while the remaining one-third is doubly coordinated by bridging dangling-O sites and dangling-H sites (Fig. 3b). (A dangling-H site denotes a water hydrogen atom not bonded to another H_2O molecule.) At the higher coverage, the adsorbate molecules are still preferentially singly or doubly coordinated. However, doubly coordinated configurations now also arise from

self-solvation, with one adsorbate molecule forming a hydrogen bond to another, to give a 'dangling-O...HCl...HCl' configuration. Past studies^{1-4,16} suggest that a minimum of three hydrogen bonds are needed to facilitate HCl ionization—one to the H atom, two to the chlorine atom. Our simulations suggest that only a few HCl molecules acquire three bonds. The formation of such a molecular adsorbate layer at low temperature reflects the paucity of surface water molecules with dangling-O and H atoms available for acid solvation.

Our Monte Carlo simulations suggest assignment of the two spectral bands observed at 50 K to singly and doubly coordinated acid molecules, respectively. For the *ab initio* model in Fig. 3a, where HCl forms a single hydrogen bond to a dangling-O atom of a water hexamer, the calculated frequency of HCl with respect to that of the gas-phase molecule is shifted by -362 cm^{-1} . This result suggests that the singly coordinated configuration gives rise to the first of the observed bands (Fig. 1). The HCl configuration in Fig. 3b, where the HCl H-atom is connected to a dangling-O of the water cluster and the Cl-atom to a dangling-H, results in substantial HCl bond stretching, but not yet in proton transfer. The calculated frequency shift of $-1,137 \text{ cm}^{-1}$ is comparable to the one of the experimentally observed shifts ($-1,175 \text{ cm}^{-1}$), indicating that the second molecular band is due to doubly coordinated HCl molecules.

In the model configuration shown in Fig. 3c, HCl replaces one of the three-coordinated water molecules in a low-energy water heptamer. After replacement and during the course of minimization, the three-coordinated HCl was found to dissociate directly, thus forming a clearly recognizable solvated H_3O^+ and Cl^- ion pair. This result is in accord with the notion that a minimum of three bonds is needed for HCl ionization^{1-4,16}. Our Monte Carlo simulations suggest that self-solvation of HCl by other HCl molecules is also significant, and another, similar model configuration was therefore constructed by replacing a second H_2O molecule of the heptamer cluster by HCl. As illustrated in Fig. 3d, this configuration

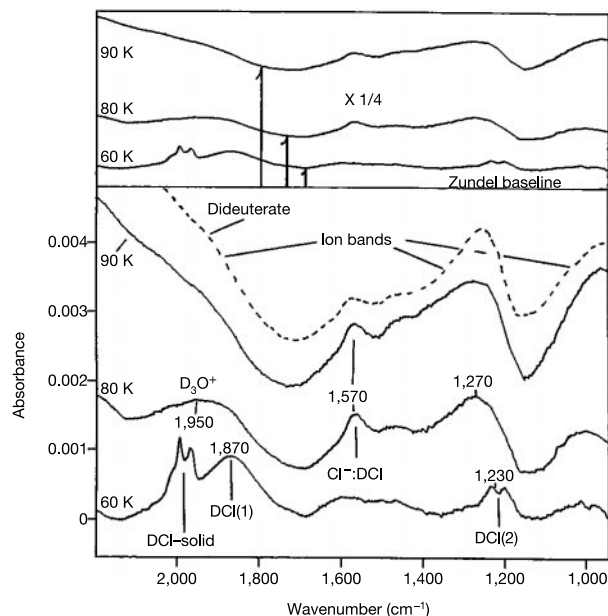


Figure 2 FTIR difference spectra for DCl adsorbate on D_2O ice at $\sim 30\%$ cover. The difference spectra are shown with (top panel) and without (bottom panel) the underlying continuum. The spectrum at 60 K in the bottom panel displays two molecular adsorbate features, DCl(1) and DCl(2), and the residual DCl nanoparticle band. The apparent doublet structure of the DCl(2) band reflects bending motion of surface water centred near $1,213 \text{ cm}^{-1}$. Ion bands are dominant at higher temperatures. At 90 K, the surface spectrum approaches that of an ionic amorphous dideuterate film (illustrated by the dashed line). In the top panel, arrows mark the growth of the Zundel continuum for the range of temperatures; the baseline was extended from above $3,000 \text{ cm}^{-1}$ (ref. 19). The spectra of the bottom panel do not reflect the continuum as they are offset for clarity. The intense continuum at 90 K reflects the abundance of D_5O_2^+ cations. Owing to acid self-solvation at the relatively high coverage used, $\sim 15\%$ ionization is obtained already at 60 K.

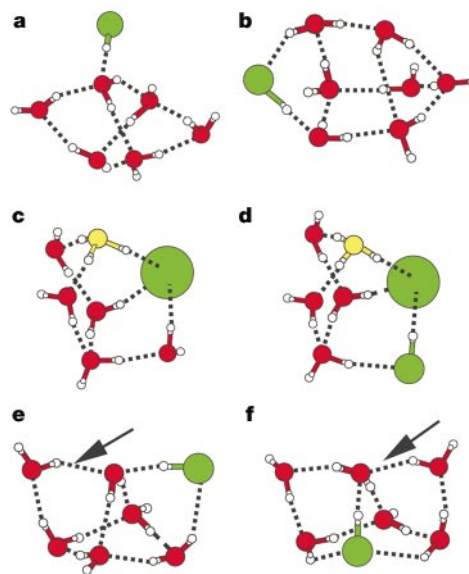


Figure 3 Cluster models used in *ab initio* calculations. The configurations and the normal frequencies were obtained on the second-order Moller-Plesset perturbation theory level in the augmented correlation consistent double zeta basis²⁷, using the Gaussian 98 program²⁸. H is shown white, Cl green, O in H_2O red, and O in H_3O^+ yellow. Hydrogen bonds are indicated by dashed lines; those marked by arrows impede HCl solvation. The H...Cl bond lengths (in Å) are: **a**, 1.31; **b**, 1.38; **c**, 1.80; **d**, 1.84 and 1.36; **e**, 1.33; and **f**, 1.34. The HCl frequency shifts cm^{-1} with respect to the gas phase are: **a**, -362 ; **b**, $-1,137$; **e**, -615 ; and **f**, -749 .

also leads to an ionized structure, with one HCl and one H₂O forming hydrogen bonds to Cl⁻. Further *ab initio* studies (results not shown) on the cyclic clusters (HCl)(H₂O)_{n=3,4} and (HCl)₂(H₂O)_{n=2,3} show comparable stretching of molecular HCl by hydrogen-bonding of either H₂O or another HCl to the Cl atom. These studies also indicate that HCl molecules, bonded to strongly stretched HCl, contribute to the first adsorption band of the observed spectra.

The evolution of the experimental spectrum with increasing temperature is shown in Fig. 2 for the deuterated system. As the temperature is raised to 80 K, new bands are observed at 1,950 and 1,570 cm⁻¹. Based on normal mode analysis of the models shown in Fig. 3c and d, these bands are assigned respectively to the OD stretch of D₃O⁺ in contact with Cl⁻, and to the stretch of the DCl molecule hydrogen-bonded to Cl⁻. (These bands appear also in spectra of the amorphous monodeuterate solid, which is rich in contact ions and contains some molecular HCl.) The increase of the contact-ion signal in going from the 60 K spectrum to the 80 K spectrum indicates that an increasing number of DCl molecules acquire a third hydrogen bond. This might occur through thermal cleaving of strained hydrogen bonds on the ice surface, and/or through DCl self-solvation following adsorbate diffusion.

Figure 3e and f show two *ab initio* cluster models containing a two- and three-coordinated HCl molecule, respectively, which is only moderately stretched. The structures illustrate that effective hydrogen bonding, capable of inducing ionization, requires HCl bonding to a two-coordinated dangling-O molecule, that is, one that does not accept a proton from another H₂O molecule. This finding is in accord with theoretical studies of proton transfer in liquid H₂O (refs 24, 25). The experimental observation of strongly stretched HCl already at 50 K (Fig. 1) suggests that breaking of such bonds requires little activation energy, and that it is promoted by the strengthening of the remaining hydrogen bonds on HCl polarization.

As the temperature is raised, the adsorbate spectra display growth of the intense underlying continuum absorption due to the Zundel ion¹⁹ D₂O-D⁺-OD₂ (Fig. 2, top). At coverages exceeding 20%, the surface layer at 90–95 K approaches the composition of the amorphous ionic dideuterate (dihydrate), as seen from the good match to spectroscopic bands of a film prepared with a ratio of 1:2 acid to water (Fig. 2, dashed curve). The composition of the crystalline dihydrate analogue was identified by X-ray diffraction²⁶ as H₅O₂⁺Cl⁻. Formation of the Zundel-ion-rich surface layer thus constitutes evidence for proton transfer away from Cl⁻. □

Received 13 November 2001; accepted 8 April 2002.

- Packer, M. J. & Clary, D. C. Interaction of HCl with water clusters: (H₂O)_nHCl, n = 1–3. *J. Phys. Chem.* **99**, 14323–14333 (1995).
- Re, S., Osamura, Y., Suzuki, Y. & Schaefer, H. F. Structures and stability of hydrated clusters of hydrogen chloride, HCl(H₂O)_n, n = 1–5. *J. Chem. Phys.* **109**, 973–977 (1998).
- Bacelo, D. E., Binning, R. C. & Ishikawa, Y. *Ab initio* Monte Carlo simulated annealing study of HCl(H₂O)_n (n = 3, 4) clusters. *J. Phys. Chem. A* **103**, 4631–4640 (1999).
- Chaban, G. M., Gerber, R. B. & Janda, K. C. Transition from hydrogen bonding to ionization in (HCl)_n(NH₃)_n and (HCl)_n(H₂O)_n clusters. *J. Phys. Chem. A* **105**, 8323–8332 (2000).
- Hanson, D. R. & Ravishankara, A. R. Investigation of the reactive and non-reactive processes involving ClONO₂ and HCl on water and nitric acid doped ice. *J. Phys. Chem.* **96**, 2682–2691 (1992).
- Horn, A. B., Chesters, M. A., McCoustra, M. R. S. & Sodeau, J. R. Adsorption of stratospherically important molecules on thin D₂O films using reflection absorption infrared spectroscopy. *J. Chem. Soc. Faraday Trans.* **88**, 1077–1078 (1992).
- Chu, L. T., Leu, M. T. & Keyser, L. F. Uptake of HCl in water ice and nitric acid ice films. *J. Phys. Chem.* **97**, 7779–7785 (1993).
- Delzeit, L., Rowland, B. & Devlin, J. P. Infrared-spectra of HCl complexed/ionized in amorphous hydrates and at ice surfaces in the 15–90 K range. *J. Phys. Chem.* **97**, 10312–10318 (1993).
- Robertson, S. H. & Clary, D. C. Solvation of hydrogen halides on the surface of ice. *Faraday Discuss.* **100**, 309–320 (1995).
- Banham, S. E., Horn, A. B., Koch, T. G. & Sodeau, J. R. Ionisation and solvation of stratospherically relevant molecules on ice films. *Faraday Discuss.* **100**, 321–332 (1995).
- Graham, J. D. & Roberts, J. T. Interaction of HCl with crystalline and amorphous ice—implications for the mechanisms of ice-catalyzed reactions. *Geophys. Res. Lett.* **22**, 251–254 (1995).
- Gertner, B. J. & Hynes, J. T. Model molecular dynamics simulation of hydrochloric acid ionization at the surface of stratospheric ice. *Faraday Discuss.* **110**, 301–322 (1998).

- Isakson, M. J. & Sitz, G. O. Adsorption and desorption of HCl on ice. *J. Phys. Chem. A* **103**, 2044–2049 (1999).
- Kang, H., Shin, T. H., Park, S. C., Kim, I. K. & Han, S. J. Acidity of hydrogen chloride on ice. *J. Am. Chem. Soc.* **122**, 9842–9843 (2000).
- Allouche, A., Couturier-Tamburelli, I. & Chiavassa, T. *Ab initio* model study of the mechanism of hydrogen chloride ionization on ice. *J. Phys. Chem. B* **104**, 1497–1506 (2000).
- Svanberg, M., Pettersson, J. B. C. & Bolton, K. J. Coupled QM/MM molecular dynamics simulations of HCl interacting with ice surfaces and water clusters—Evidence of rapid ionization. *J. Phys. Chem. A* **104**, 5787–5798 (2000).
- Mantz, Y. A. *et al.* The interaction of HCl with the (0001) face of hexagonal ice studied theoretically via Car-Parrinello molecular dynamics. *Chem. Phys. Lett.* **348**, 285–292 (2001).
- Devlin, J. P. & Buch, V. Surface of ice as viewed from combined spectroscopic and computer modeling studies. *J. Phys. Chem.* **99**, 16534–16548 (1995).
- Zundel, G. Hydrogen bonds with large proton polarizability and proton transfer processes in electrochemistry and biology. *Adv. Chem. Phys.* **111**, 1–217 (2000).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Votava, C., Ahlrichs, R. & Geiger, A. The HCl-HCl interaction: From quantum mechanical calculations to properties of the liquid. *J. Chem. Phys.* **78**, 6841–6848 (1983).
- Uras, N., Buch, V. & Devlin, J. P. Hydrogen bond surface chemistry: interaction of NH₃ with an ice particle. *J. Phys. Chem. B* **104**, 9203–9209 (2000).
- Devlin, J. P., Joyce, C. & Buch, V. Infrared spectra and structures of large water clusters. *J. Phys. Chem. A* **104**, 1974–1977 (2000).
- Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462 (1995).
- Ando, K. & Hynes, J. T. Molecular mechanism of HCl acid ionization in water. *J. Phys. Chem. B* **101**, 10464–10478 (1997).
- Lundgren, J. O. & Olovsson, I. The crystal structure of hydrogen chloride dihydrate. *Acta Crystallogr.* **23**, 966–971 (1967).
- Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. *J. Chem. Phys.* **90**, 1007–1023 (1989).
- Frisch, M. J. *et al.* *Gaussian 98* (Gaussian, Pittsburgh, 1998).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.B. (e-mail: viki@fh.huji.ac.il).

A palaeontological solution to the arthropod head problem

Graham E. Budd

Department of Earth Sciences (Historical Geology and Palaeontology), Uppsala University, Norbyvägen 22, Uppsala, Sweden SE-752 36

The composition of the arthropod head has been one of the most controversial topics in zoology, with a large number of theories being proposed to account for it over the last century¹. Although fossils have been recognized as being of potential importance in resolving the issue^{2,3}, a lack of consensus over their systematics^{4,5} has obscured their contribution. Here, I show that a group of previously problematic Cambrian arthropods from the Burgess Shale and Chengjiang faunas form a clade close to crown-group euarthropods, the group containing myriapods, chelicerates, insects and crustaceans⁶. They are characterized by modified or even absent endopods, and two pre-oral appendages. Comparison with reconstructions of the crown-group euarthropod ground plan⁶ and recent investigations into onychophorans^{7,8} demonstrates that these two appendages are the first antenna (of extant crustaceans) and a more anterior appendage associated with an ocular segment. The latter appendage has been reduced in all crown-group euarthropods. Its most likely relic is as a component of the labrum⁹. These fossils thus tie together results from disparate living groups (onychophorans and euarthropods).

One of the main difficulties in unravelling the complexities of the arthropod head has been the assumed loss of appendages and reorganization and reduction of putative head segments, especially

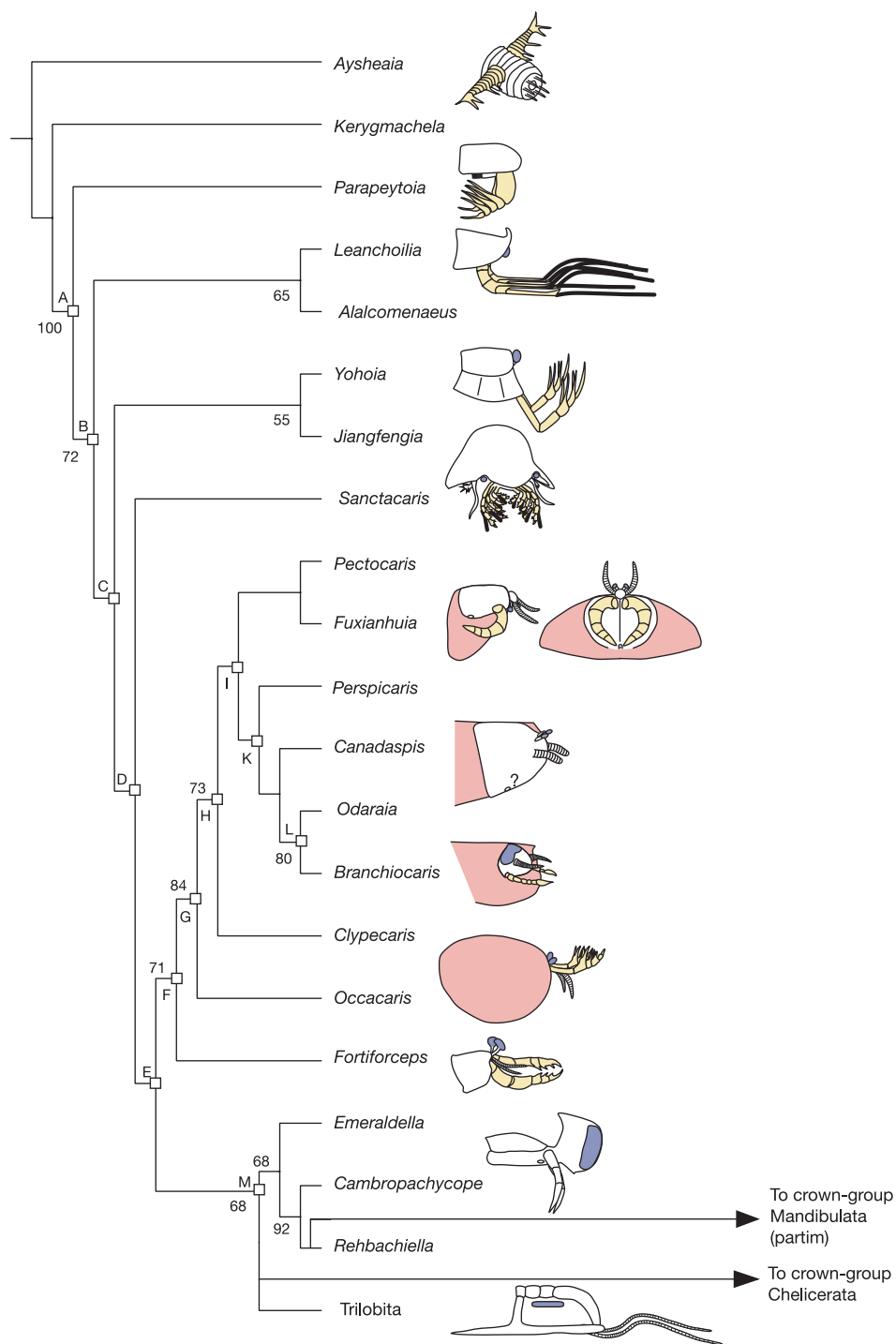


Figure 1 Cladistic analysis (see Supplementary Information for details) of upper stem-group euarthropods, with schematic drawings of the pre-oral region of selected taxa. Frontal appendage shown in yellow; A1 in white; free carapace in pink; eyes in purple. Putative branching positions of chelicerates¹² and mandibulates^{11,18} are shown. One of two most-parsimonious trees (branch-and-bound search); tree lengths 62; rescaled consistency index 0.530. Bootstrap supports over 50% based on 500 replicates shown at nodes. *Aysheaia* was used as an outgroup. Apomorphies for selected nodes (optimized with accelerated transformation) as follows. A: Ventral mouth; frontal appendage with few podomeres, grasping structures terminally; well-articulated limbs; headshield; gnathobasic limbs; biramous limbs; two post-oral cephalic appendages. B: Loss of 'peytoia' mouthparts; distinct tergites; limb outer branched with long setae. C: Three post-

oral cephalic appendages. D: A1 differentiated as antenniform appendage. E: A1 uniramous. F: Two pre-oral appendages developed; multi-segmented inner limb branch; outer branch simple oval flap with small fringing setae (perhaps constituted with ovate lobes as in *Fortiforceps* and *Canadaspis*); loss of gnathobases. G: Head shield developed as free carapace extending posteriorly over anterior thoracic appendages; headshield bivalved; loss of abdominal tergite pleurae; ?loss of post-oral cephalic appendages (uncertain in most taxa). H: Frontal appendage sub-chelate, ventral; no posterior limbs; terminal furca. I: Straight hinge. K: Posterior spine on valve. L: High number of segments (>30); small posterior limbs; outer limb branch small, triangular flap. M: Loss of frontal appendage; outer limb branch with more than one podomere. In-group crown-group euarthropod relationships are not well resolved in this analysis.

in the pre-oral region. However, Cambrian taxa have been suggested to have a less complex organization, and may thus provide vital clues to the evolution of the arthropod head². Unfortunately, the many euarthropod taxa (that is, the least inclusive clade to encompass chelicerates, crustaceans, insects, myriapods and pycnogonids⁶) known from the various exceptionally well-preserved Cambrian faunas have long been controversial^{4,5}. Early views of them as 'primitive' members of extant classes were replaced by the idea that these euarthropod taxa represent the end-points of many independent lines of evolution¹⁰. More recently, cladistically based investigations have clarified the stem-lineage of the crustaceans¹¹, and begun to unravel the complexities of the many putative stem-group chelicerates¹². This has allowed the reconstruction of the basal euarthropod head to possess a single pair of pre-oral appendages, corresponding to the antennae of trilobites and first antenna (A1) of crustaceans⁶, myriapods and insects, but of disputed homology with the chelicera of chelicerates. However, the link between these derived taxa—all probable crown-group euarthropods—and lobopod-like putative stem-group euarthropods such as *Kerygmachela*¹³ has remained obscure; yet it is a critical link if the origins of the major features shared by living euarthropods are to be traced.

To resolve this outstanding issue, arthropod taxa from the major Cambrian lagerstätten (the Lower–Middle Cambrian Sirius Passet, Chengjiang and Burgess Shale faunas and the Upper Cambrian Orsten fauna) were coded and analysed with PAUP³¹. Twenty-one well-preserved taxa were analysed, with 33 characters; the lobopod *Aysheaia*¹⁴ was used as an outgroup. Inclusion of living taxa would in general raise considerable complications in terms of character inapplicability¹⁵ and coding, and enormously expand the scope of the analysis. However, putative proxies for stem-group morphologies may be used to replace the crown groups. Putative representatives of the mandibulate (*Cambropachycope*¹¹) and chelicerate (Trilobita¹², *Emeraldella*¹⁶) stem groups were thus included. *Rehbachella*¹⁷, a crown-group crustacean, was also included, as it possesses differentiated first (A1) and second (A2) antennae—a unique combination in the extant fauna—and has been shown to share important synapomorphies with the crustacean stem-lineage⁶, allowing sensible character coding. Insects are here considered as mandibulates (and are thus represented by the stem-

group 'crustacean' *Cambropachycope*), as are myriapods, which, despite controversy, are allied by all morphological data with other mandibulates¹⁸, a position that is also supported by some molecular data¹⁸. Tree robustness was tested with bootstrap analysis. Two most-parsimonious trees (differing only in detail) were obtained (one tree shown in Fig. 1). Most of the arthropods plotted in this analysis fall outside the euarthropod crown-group, and are thus less derived than any living taxa. However, the most striking feature of the tree is the presence of a well-supported clade of taxa including *Fortiforceps*¹⁹ and the well-known *Canadaspis*²⁰ and *Fuxianhuia*^{3,19}. Members of this clade are in general characterized by a flap-like exopod (uniquely similar in the case of *Fortiforceps* and *Canadaspis*) with short or absent fringing setae, and modified endopods, basally possessing many podomeres, with more derived forms modifying or even losing the endopod altogether. With the exception of *Fortiforceps* (basal within the clade), these taxa also all possess a bivalved carapace, with more derived members developing a long horizontal hinge, together with high segment numbers and loss of posterior limbs. One other apparent exception is *Fuxianhuia*, which, although it possesses a posteriorly extending carapace that is free from the anterior thoracic segments, is not known to be bivalved. However, the presence of laterally preserved specimens show that the carapace did fold along a straight line during compression, and dorsoventrally flattened specimens (for example, Fig. 2a, b of ref. 3; Fig. 8b of ref. 16) also show an unusual median crease in the cephalic shield. *Fuxianhuia* seems therefore also to have been effectively bivalved.

The most important feature of this clade (and the immediately more basal *Sanctacaris*²¹) is, however, the presence of two pre-oral appendages, one antenniform, and the other a characteristic robust structure with few (usually 5) podomeres and a spinose or subchelate tip. This appendage is extremely similar to the so-called 'great appendage' possessed by (in this analysis) less-derived taxa such as *Yohoia* and *Leancoilia*². Indeed, such an appendage is also present in the anomalocaridids, with the one of *Parapeytoia*²² (a derived anomalocaridid with gnathobasic biramous limbs) being particularly similar to that of *Yohoia* and *Leancoilia*. It is also present in even more basal forms such as *Kerygmachela*¹³ (the 'frontal appendage'²), *Pambdelurion*²³ and the lobopodian *Aysheaia*, where it appears to be the most anterior appendage, just posterior to

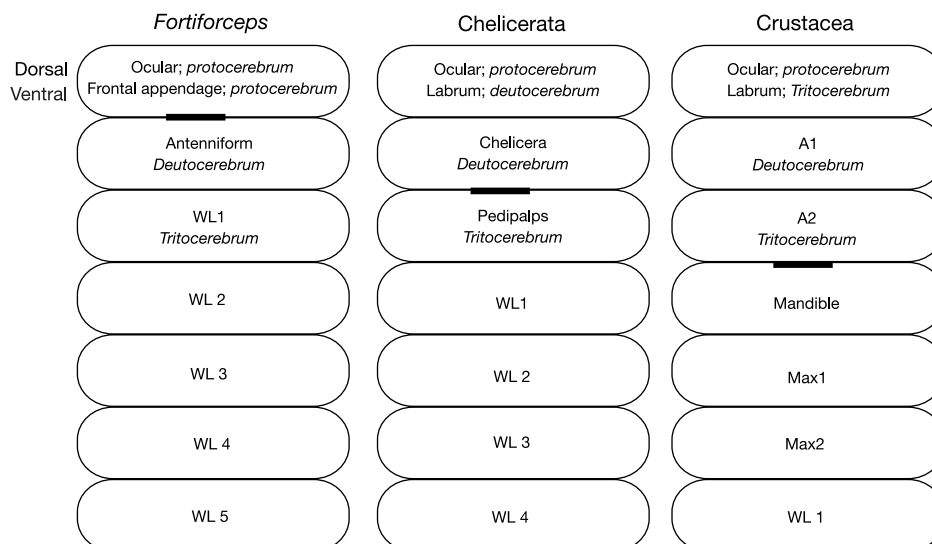


Figure 2 Scheme for arthropod head segmentation implied by this analysis. The homologies between a stem-group euarthropod (*Fortiforceps*) and the crustacean/chelicerate clades are shown. Position of mouth in adult shown as bold line. An implied innervation for each structure is shown in italics. I note the complex history of innervation

implied for originally protocerebral structures, related to ventral rotation of the mouth (in broad agreement with ref. 2). A1, first antenna; A2, second antenna; Max1, first maxilla; Max2, second maxilla; WL, walking limb.

an anterior mouth. This frontal appendage therefore appears to be a homologous feature that appears deep in arthropod phylogeny².

The clade consisting of *Sanctacaris* and all more derived taxa, conversely, is primitively characterized by the presence of an antenniform appendage that in origin seems to be posterior to the frontal appendage. In relatively derived taxa (for example, *Fuxianhuia*) the antenniform appendage actually lies anterior to the frontal appendage, but this placement results from ventral rotation of the frontal appendage to follow the mouth^{13,24}. The antenniform appendage is therefore relatively dorsal, and its anterior–posterior alignment with respect to the frontal appendage is a matter of the degree of rotation. Moreover, the basal condition—deduced from *Occacaris* (Figs 1.1 and 2 of ref. 25) and *Fortiforceps* (see Figs 32A and 33B of ref. 19)—of the relevant clade is for both appendages apparently to insert very closely longitudinally, so that, primitively for this clade, the frontal appendage may in fact be the more anterior. The antenniform appendage is retained by the crownward taxa such as *Emeraldella* and crown-group euarthropods⁶. The evolutionary history of the pre-oral region of euarthropods therefore appears to be relatively stable. Low in the euarthropod stem group, the head is characterized by a single pair of ‘frontal appendages’ that were originally flanking a terminal mouth. These appendages are retained in more sclerotized forms such as the anomalocaridids, now showing a characteristic reduction of podomeres and spines, and into taxa that fit more readily into conventional categories of ‘arthropods’ such as *Yohoia* and *Leancoilia*. Higher euarthropods are all characterized by the presence of antennae, which persist into the crown group. However, within and just outside the new clade identified here, a critical juxtaposition of the frontal and antenniform appendages is seen, notably in *Fuxianhuia* (see Figs 8 and 9E of ref. 19), *Fortiforceps*, *Occacaris* and *Sanctacaris* (plate 1 and text-fig. 1 of ref. 21). This juxtaposition demonstrates that the two appendages are not homologous; nor are either homologous with the eye stalk. Furthermore, as the frontal appendage is the most anterior appendage (and the only differentiated one) in lower stem-group taxa, the antenniform appendage must be derived from behind it. Taken together, this evidence demonstrates that the lack of the frontal appendage in crown-group euarthropods implies a loss rather than transformation into another well-developed appendage.

What are the identities of the frontal and antenniform appendages? Recent results from the onychophorans^{7,8}, which lie basal to the entire clade, demonstrate that the onychophoran ‘antenna’ (likely to be homologous with the frontal appendage of *Aysheaia* and the stem-group euarthropods) is innervated no more posteriorly than the eye. This implies that the frontal appendage is primitively innervated from the protocerebrum. Further, the present analysis supports the notion that basal crown-group euarthropods possessed only a single pre-oral, antenniform appendage^{6,11}. The frontal appendage lying in front of this appendage must therefore belong to a segment that does not possess a well-developed appendage in any extant euarthropod, (supporting one possibility suggested by ref. 2). The fossil record thus strongly supports the existence of an ‘ocular segment’ that once bore an appendage in basal euarthropods^{2,3}. The possible homology of this structure with any crown-group euarthropod structure is unclear, but given that the frontal appendage becomes ventralized, and is essentially a feeding organ^{24,26}, one strong possibility is that the labrum *sensu lato* may partly incorporate it (together with the hypostome). This analysis is thus consistent with molecular results that have been taken to demonstrate that the labrum is appendiculate⁹, and with morphological results suggesting it is derived from the anterior of the euarthropod head²⁷. The likelihood that the labrum is an endite of the ‘intercalary segment’⁹ however, seems low, because crustaceans have both a fully developed second antenna derived from this segment and a labrum.

Structures derived from the ocular segment therefore today have

a complex distribution, which may contribute to the difficulty in identifying it. The scheme implied for the euarthropod head is given in Fig. 2. It should be noted that in this scheme, there is no space for a pre-oral and nonsegmental ‘acron’²⁷. As the mouth started terminal within the clade^{13,24} (thus supporting the Ecdysozoa²⁸ concept), any deep pre-oral structure would have to have migrated behind the mouth before the origin of the Ecdysozoa, and then remigrated anteriorly again when the mouth became ventralized. Such a structure has yet to be identified.

This analysis brings fresh insight into, but does not resolve, the continuing problem presented by the chelicerate chelicera. If this appendage developed from the antenna of putative stem-group chelicerates such as the trilobites (favoured here), then it is homologous to A1 (as suggested by molecular results²⁹); but if it developed from the appendage behind the antenna, then it is homologous to the second antenna (the result of classical morphology³⁰). In the former case, this work suggests one pre-chelicerate segment; in the latter, two. In theory, the fossil record could resolve this issue by revealing clear stem-group chelicerates with both an antenna (equivalent to A1) and a more posterior appendage (equivalent to chelicera) differentiated. The mere presence of two pre-oral appendages, one of which is chelicerate, in a fossil arthropod form would not, however, demonstrate that the chelicera is homologous to the second antenna (A2), because these two appendages could be the frontal appendage and A1. □

Received 16 October 2001; accepted 18 February 2002.

1. Rempel, J. G. The evolution of the insect head: the endless dispute. *Q. Entomol.* **11**, 7–25 (1975).
2. Dewel, R. A. & Dewel, W. C. The place of tardigrades in arthropod evolution. *Syst. Assoc. Sp. Vol.* **55**, 109–123 (1997).
3. Chen, J. Y., Edgecombe, G. D., Ramsköld, L. & Zhou, G. Head segmentation in Early Cambrian *Fuxianhuia*: implications for arthropod evolution. *Science* **268**, 1339–1343 (1995).
4. Wills, M. A., Briggs, D. E. G., Fortey, R. A. & Wilkinson, M. The significance of fossils in understanding arthropod evolution. *Verh. Dcut. Zool. Gessell.* **88**, 203–215 (1995).
5. Simonetta, A. M. & Della Cave, L. in *The Early Evolution of Metazoa and the Significance of Problematic Taxa* (eds Simonetta, A. M. & Conway Morris, S.) 189–244 (Cambridge Univ. Press, Cambridge, 1991).
6. Walossek, D. in *Proc. 4th Int. Crustacean Congr.* (eds Schram, F. R. & von Vaupel Klein, J. C.) 3–27 (Brill Academic, Leiden, 1999).
7. Eriksson, B. J. & Budd, G. E. The cephalic nerves of the Onychophora and their bearing on our understanding of head segmentation of stem-group evolution of Arthropoda. *Arth. Struct. Dev.* **29**, 197–209 (2001).
8. Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran *Euperipatoides kanangensis* with particular reference to the central nervous system. *J. Morphol.* (in the press).
9. Haas, M. S., Brown, S. J. & Beeman, R. W. Pondering the procephalon: the segmental origin of the labrum. *Dev. Genes. Evol.* **211**, 89–95 (2001).
10. Whittington, H. B. Early arthropods, their appendages and relationships. *Syst. Ass. Spec.* **12**, 253–268 (1979).
11. Walossek, D. & Müller, K. J. Upper Cambrian stem-lineage crustaceans and their bearing upon the monophyletic origin of Crustacea and the position of *Agnostus*. *Lethaia* **23**, 409–427 (1990).
12. Edgecombe, G. D. & Ramsköld, L. Relationships of Cambrian Arachnata and the systematic position of Trilobita. *J. Paleontol.* **73**, 263–287 (1999).
13. Budd, G. E. The morphology and evolutionary significance of *Kerygmachela kierkegaardi* Budd (Buen Formation, Lower Cambrian, North Greenland). *Trans. R. Soc. Edinb. Earth Sci.* **89**, 249–290 (1999).
14. Whittington, H. B. The lobopodian animal *Aysheaia pedunculata* Walcott, Middle Cambrian, Burgess Shale, British Columbia. *Phil. Trans. R. Soc. Lond. B* **284**, 165–197 (1978).
15. Richter, S. & Scholtz, G. Phylogenetic analysis of the Malacostraca (Crustacea). *J. Zool. Syst. Evol. Res.* **39**, 113–136 (2001).
16. Bruton, D. L. & Whittington, H. B. *Emeraldella* and *Leancoilia*, two arthropods from the Burgess Shale, Middle Cambrian, British Columbia. *Phil. Trans. R. Soc. Lond. B* **300**, 553–585 (1983).
17. Walossek, D. The Upper Cambrian *Rehbachella* and the phylogeny of Branchiopoda and Crustacea. *Fossils Strata* **32**, 1–202 (1993).
18. Giribet, G., Edgecombe, G. D. & Wheeler, W. C. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).
19. Hou, X.-G. & Bergström, J. Arthropods of the Lower Cambrian Chengjiang fauna, southwest China. *Fossils Strata* **45**, 1–116 (1997).
20. Briggs, D. E. G. The morphology, mode of life and affinities of *Canadaspis perfecta* (Crustacea: Phyllocarida), Middle Cambrian, Burgess Shale, British Columbia. *Phil. Trans. R. Soc. Lond. B* **281**, 439–487 (1978).
21. Briggs, D. E. G. & Collins, D. A Middle Cambrian chelicerate from Mount Stephen, British Columbia (Canada). *Palaeontology* **31**, 779–798 (1988).
22. Hou, X.-G., Bergström, J. & Ahlberg, P. *Anomalocaridids* and other large animals in the Lower Cambrian Chengjiang fauna of southwest China. *GFF* **117**, 3–183 (1995).
23. Budd, G. E. Stem-group arthropods from the Lower Cambrian Sirius Passet fauna of North Greenland. *Syst. Assoc. Sp. Vol.* **55**, 125–138 (1997).
24. Dewel, R. A., Budd, G. E., Castano, D. F. & Dewel, W. C. The organization of the subesophageal nervous system in tardigrades: insights into the evolution of the arthropod hypostome and tritocerebrum. *Zoologischer Anzeiger* **238**, 191–203 (1999).

25. Hou, X.-G. New rare bivalved arthropods from the Lower Cambrian Chengjiang fauna, Yunnan, China. *J. Palaeontol.* **73**, 102–116 (1999).
26. Bousfield, E. L. A contribution to the natural classification of Lower and Middle Cambrian arthropods: food-gathering and feeling mechanisms. *Amphipacifica* **2**, 3–33 (1995).
27. Scholtz, G. Evolution of developmental patterns in arthropods—the analysis of gene expression and its bearing on morphology and phylogenetics. *Zoology* **103**, 99–111 (2001).
28. Aguinaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
29. Telford, M. J. & Thomas, R. A. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* **95**, 10671–10675 (1998).
30. Winter, G. Beiträge zur Morphologie und embryologie des vorderen Körperabschnitts (Cephalosoma) der Pantopoda Gerstaecker, 1863. *Z. Zool. Syst. Evol.* **18**, 27–61 (1980).
31. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1* (Illinois Natural History Survey, Champaign, Illinois, 1993).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This work grew out of a collaboration with R. A. Dewel, whom I thank for discussions that helped shape many of the ideas contained herein. D. Waloszek and J. Eriksson have also provided much help and insight. I also thank J. Bergström, H. Xianguang, D. Collins and D. Erwin for allowing access to material. The comments of G. Edgecombe substantially improved the paper. This work was funded by the Swedish Research Council (VR).

Competing interests statement

The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: graham.budd@pal.uu.se).

Climatic influence on a marine fish assemblage

Martin J. Attrill* & Michael Power†

* Department of Biological Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK

† Department of Biology, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Understanding the fluctuations in marine fish stocks is important for the management of fisheries, and attempts have been made to demonstrate links with oceanographic and climatic variability^{1–3}, including the North Atlantic Oscillation (NAO)^{4,5}. The NAO has been correlated with a range of long-term ecological measures^{6,7}, including certain fish stocks^{8,9}. Such environmental influences are most likely to affect susceptible juveniles¹⁰ during estuarine residency, as estuaries are critical juvenile nursery or over-wintering habitats¹¹. Here we show that, during a 16-year period, climatic forcing (by means of the NAO) is consistently the most important parameter explaining variation in assemblage composition, abundance and growth of juvenile marine fish during estuarine residency. A possible mechanism for the effect of the NAO is a temperature differential between estuarine and marine waters that allows fish to facultatively exploit optimal thermal habitats. The connection has potentially important implications for the size and numbers of individuals recruited to the fishery, for understanding and predicting the composition of juvenile fish stocks using estuaries, and for the appropriate conservation of estuarine systems in relation to fish stocks.

Over the past century, the principal issue in fishery science has arguably been understanding recruitment variability^{12,13}, either as a tool for predicting future success of a fishery or as a method for explaining variation in fish stocks. Traditionally, investigations

correlating fish abundance and growth to environmental variables have focussed on the role of marine temperatures for commercially important species such as cod (*Gadus morhua*)^{3,14}. Similarly, stock recruitment levels have been associated with variations in marine temperature during the first years of life^{9,15}. In the North Atlantic, sea surface temperature (SST) has been related to the dominant signal of climatic variability¹⁶, the NAO, which has a potentially influential role in determining stock recruitment levels^{3,8}. However, many commercially important marine fish spend critical juvenile stages in estuarine nursery grounds¹¹, which may act as thermal buffers against more severe open-sea conditions, and therefore are not affected directly by marine conditions.

To investigate the potential impact of climatic variation on juvenile fish stages, we have analysed one of the most comprehensive long-term data sets available for estuarine fish communities, the 16-year data set compiled from the Thames estuary, UK. This has been used previously to model intra-annual responses of fish and invertebrate populations to physico-chemical trends^{17–19}. As the fish assemblage is primarily composed of juvenile marine species using the estuary as a nursery²⁰, we have collated the data by year to facilitate the investigation of interannual relationships between environmental variables and fluctuations in fish assemblage composition, abundance and juvenile growth. The environmental variables included measurements of estuarine water quality, freshwater flow, North Sea SST, and climatic variation as defined by the North Atlantic Oscillation Index (NAOI). The species selected *a priori* for population analysis included all commercially important marine groups (flatfish (*Solea solea*, *Platichthys flesus*, *Pleuronectes platessa*, *Limanda limanda*), gadoids (*Merlangius merlangus*, *Trisopterus luscus*, *T. minutus*), clupeids (*Clupea harengus*, *Sprattus sprattus*) and bass (*Dicentrarchus labrax*)), the dominant estuarine fish species (gobies (*Pomatoschistus* spp.), eel (*Anguilla anguilla*), smelt (*Osmerus eperlaunus*), pogge (*Agonus cataphractus*) and Nilsson's pipefish (*Syngnathus rostellatus*)), and invertebrates important either as prey or as predators¹⁸ (brown shrimp (*Crangon crangon*) and gelatinous species (*Pleurobrachia pileus*, *Aurelia aurita*)). Where possible, growth during estuarine residency was examined in relation to the NAOI, as were standard summary measures (for example, diversity, evenness) of assemblage composition.

From a total of 50 possible relationships between biological measures and the NAOI, significant relationships were apparent in 24 cases (Fig. 1 and Table 1; see also Supplementary Information). All fish assemblage measures demonstrated significant relationships with NAOI, and when compared with other environmental data NAOI was always the best predictor (highest r^2 value) of assemblage variation. Similarly, where significant relationships between fish population or growth measures and environmental variables were apparent, NAOI or NAOI that was lagged by 1 year best explained the variation in all but two cases (Gobiidae abundance and bass average size). Nonsignificant ($P > 0.05$) or statistically inadequate regression models (autocorrelation, non-normal residuals) relating abundance to time suggested that temporal trends in the data did not account for observed correlations with the NAOI. Growth and abundance were not significantly correlated ($P > 0.05$), suggesting that density dependence was not a factor in mediating observed correlations between NAOI and growth. Of the invertebrates, two species demonstrated significant abundance relationships: *Aurelia* and *Crangon*. Table 1 presents the statistically significant relationships with NAOI, which explained up to 54% of the variation over 16 years in assemblage measures, 69% of fish abundance, 76% of fish growth and 46% of invertebrate abundance.

To control further for possible temporal or density-dependent effects on correlations with the NAOI, partial correlation coefficients for abundance and growth data and the NAOI that controlled for temporal trends and abundance, were computed. These supported the regression results (Table 2; see also Supplementary

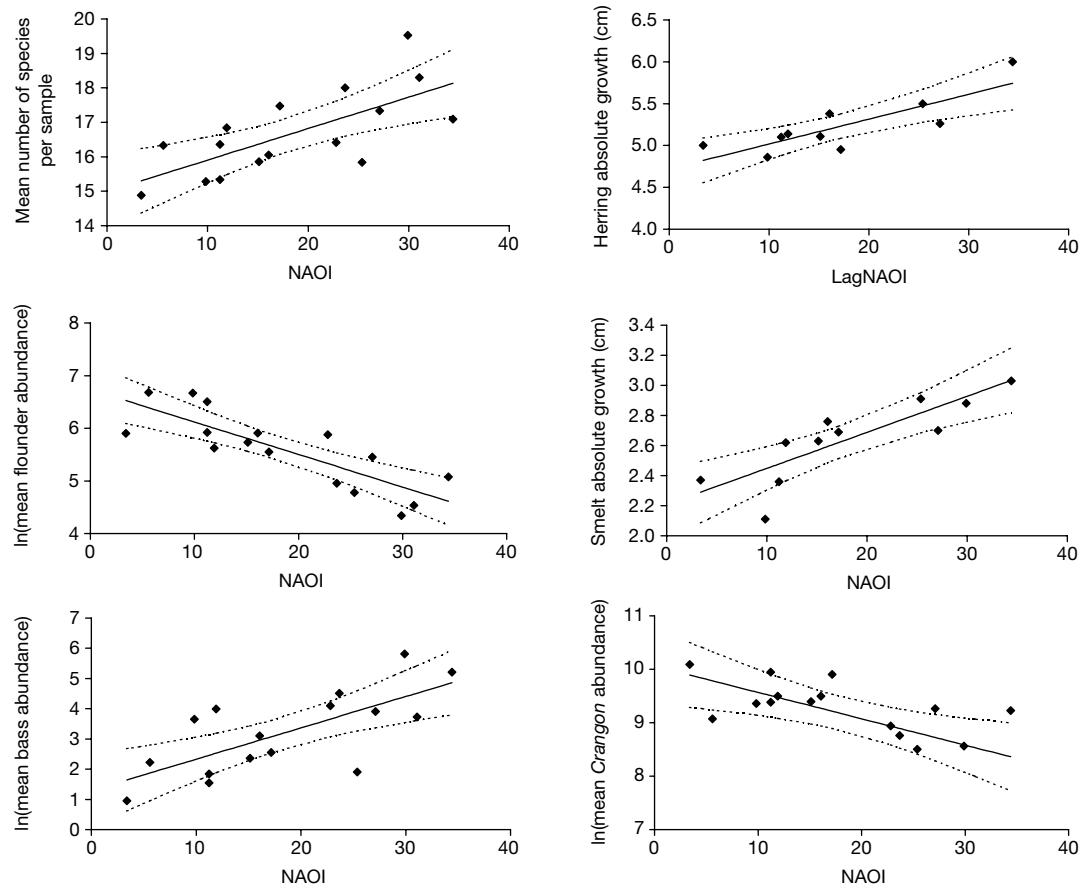


Figure 1 Examples of relationships between NAOI and fish assemblage, population and growth measures. For all cases, NAOI was either the only variable significantly related with the dependent variable or had the highest r^2 value. Regression details are in Table 1.

Solid line, model; dashed line, 95% confidence intervals; filled symbols, data points; LagNAOI, NAOI with a lag period of one year.

Table 1 Statistically acceptable linear regressions for association between NAOI or LagNAOI and biological measures

Parameter	<i>n</i>	<i>x</i>	Regression equation	r^2	<i>P</i>
Fish assemblage measures					
Mean species number per sample	16	NAOI	$y = 14.99 + 0.091x$	<u>0.493</u>	0.002
Mean Shannon–Wiener index	16	NAOI	$y = 1.287 + 0.013x$	<u>0.541</u>	0.001
Mean Pielou's evenness index	16	NAOI	$y = 0.478 + 0.004x$	<u>0.493</u>	0.002
Mean Berger–Parker dominance index	16	NAOI	$y = 0.581 - 0.005x$	<u>0.371</u>	0.012
Number of rare fish species year ⁻¹	16	NAOI	$y = 5.533 + 0.262x$	0.308	0.026
ln(total marine fish abundance)	16	NAOI	$y = 10.77 - 0.016x$	0.319	0.023
ln(total marine fish abundance)	16	LagNAOI	$y = 10.79 - 0.016x$	0.357	0.015
Fish population measures					
ln(mean flatfish abundance)	16	NAOI	$y = 6.801 - 0.062x$	<u>0.685</u>	<0.001
ln(mean flounder abundance)	16	NAOI	$y = 6.745 - 0.062x$	<u>0.667</u>	<0.001
ln(mean bass abundance)	16	NAOI	$y = 1.294 + 0.104x$	<u>0.510</u>	0.002
ln(mean Gobiidae abundance)	16	NAOI	$y = 9.397 - 0.030x$	0.328	0.020
ln(mean pipefish abundance)*	16	NAOI	$y = 1.677 + 0.068x$	0.341	0.018
Fish growth measures (cm)					
Bass average age-0 size	13	NAOI	$y = 5.915 + 0.082x$	<u>0.630</u>	0.018
Bass absolute growth*	13	NAOI	$y = 0.176 + 0.069x$	<u>0.513</u>	0.006
Dab average age-0 size	10	NAOI	$y = 4.721 + 0.030x$	<u>0.684</u>	0.003
Dab entry size	10	NAOI	$y = 3.966 + 0.040x$	<u>0.652</u>	0.005
Herring absolute growth	12	LagNAOI	$y = 0.780 + 0.021x$	<u>0.578</u>	0.004
Sprat absolute growth	14	LagNAOI	$y = 0.381 + 0.021x$	<u>0.755</u>	0.001
Smelt average age-0 size	11	NAOI	$y = 6.057 + 0.093x$	<u>0.646</u>	0.003
Smelt absolute growth	11	NAOI	$y = 2.208 + 0.024x$	<u>0.702</u>	0.001
Whiting absolute growth	13	NAOI	$y = 2.153 + 0.071x$	0.430	0.015
Invertebrate measures					
Total <i>Aurelia</i> abundance*	16	LagNAOI	$y = 1702 - 57.24x$	0.275	0.037
ln(mean <i>Crangon</i> abundance)	16	NAOI	$y = 10.063 - 0.049x$	<u>0.402</u>	0.008
ln(mean <i>Crangon</i> abundance)	16	LagNAOI	$y = 10.004 - 0.046x$	<u>0.352</u>	0.015

Total abundance (*Aurelia* only) is the number of individuals caught per year; mean abundance is the mean number of individuals caught per sample (500 million litres of filtered water). Regression *P*-values are provided; all slope *P*-values are <0.05. For growth measures, $n < 16$ owing to inadequate sample sizes ($n < 25$) for growth calculations. Underlined values indicate models that remain statistically significant after Bonferroni inequality adjustment; *x*, NAOI or LagNAOI; LagNAOI, NAOI with a lag period of 1 year. Dab, *L. limanda*; flounder, *P. flesus*; whiting, *M. merlangus*.

* Relationship not significant when data de-trended (Table 2).

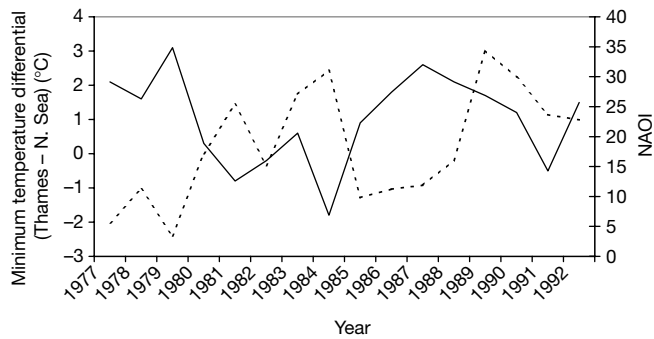


Figure 2 Differential in minimum recorded water temperatures (°C) between the Thames estuary and southern North Sea (solid line), together with NAOI (dashed line) for the period 1977–92. Temperature differential demonstrates a significant negative linear relationship with NAOI ($r^2 = 0.318$, $P = 0.023$).

Information), adding eight additional significant relationships and removing only three from those reported in Table 1. Our results suggest, therefore, that climatic variability has a principal controlling influence on the structure of the Thames fish assemblage, the growth of many resident juveniles, and the abundance of many of the dominant fish species using the estuary as a nursery area. Although the ecological effects of the NAO have been demonstrated on the abundance and growth of single species within a community⁷, or on limited community measures²¹, this study demonstrates relationships for the NAO at multiple levels of biological organization and for a wide range of species occupying a similar habitat.

All significant parameters of fish growth were positively correlated with NAOI. High NAOI values generally correspond to warm, wet and stormy winters in the northeast Atlantic⁵. As temperature strongly influences juvenile growth¹⁰, temperature has been assumed to be the causative mechanism underlying the influence of the NAO on growth⁷. However, in all cases, regressions with the

NAOI explained higher proportions of the variation in growth than either estuarine water temperatures or North Sea SST alone. Furthermore, the growth of the main species of Clupeidae was related to NAOI with a one-year lag, rather than climatic conditions during the winter of estuarine residency. This suggests that principal climate-mediated environmental influences determine growth during the earliest marine larval stage of these species, either directly or by influencing feeding conditions for the parent stock²². Accordingly, our study suggests that climate-influenced hydrographical and meteorological factors, in addition to temperature, are important in determining juvenile fish growth in estuaries. Consistently strong relationships were also apparent between measures of the assemblage composition and the NAOI, with the index alone explaining nearly 50% of the variability in the number of fish species caught per sample. Diversity increases during wet, warm winters (high NAOI), which is partly explained by the significant increase in the number of rare species (abundance less than 0.1% of total catches) sampled during high NAOI years. Most of the additional species were predominantly those with a southern distribution²³ (for example, gurnards, anchovy, wrasse, weever). The appearance of unusual species or the extension of species' ranges have been used as indicators of climate change²⁴, but the appearance of warm-water southern species may simply reflect long-term climatic cycles¹.

Most of the fish species investigated (9 out of 15) demonstrated a significant relationship between population abundance and the NAOI. The specific responses of fish populations to climatic variation can be categorized into three groups: a negative relationship with the NAOI (flatfish species, northern species²³; for example, herring (*C. harengus*)), a positive relationship with NAOI (southern species²³; for example, bass, sprat (*S. sprattus*)), or no significant relationship (for example, all gadoids, Dover sole (*S. solea*), estuarine species). As the primary use of estuaries by marine fish is as a nursery ground, the results suggest that estuarine usage, and therefore the relative importance of estuaries, is dependent on climate for many species. Increases in the population size of southern species, such as bass, during warm, high NAOI years is consistent with an opportunistic use of available thermal habitat. In

Table 2 Partial correlation coefficients measuring the association between NAOI or LagNAOI and biological measures

Parameter	n	x	r	Significance
Fish population measures				
ln(mean herring abundance)	16	LagNAOI	-0.599	* (additional relationship)
ln(mean sprat abundance)	16	NAOI	0.644	** (additional relationship)
ln(mean flatfish abundance)	16	NAOI	-0.782	**
ln(mean flounder abundance)	16	NAOI	-0.771	**
ln(mean plaice abundance)	16	NAOI	-0.596	* (additional relationship)
ln(mean Gobiidae abundance)	16	NAOI	-0.648	**
ln(mean sand goby abundance)	16	NAOI	-0.673	** (additional relationship)
ln(mean eel abundance)	16	NAOI	-0.563	* (additional relationship)
ln(mean bass abundance)	16	NAOI	0.568	*
Fish growth measures (cm)				
Whiting absolute growth	13	NAOI	0.715	**
Herring absolute growth	12	LagNAOI	0.677	**
Sprat absolute growth	14	LagNAOI	0.942	**
Dab average age-0 size	10	NAOI	0.826	**
Dab entry size	10	NAOI	0.927	**
Flounder average age-0 size	13	NAOI	0.678	** (additional relationship)
Plaice average age-0 size	10	NAOI	0.762	** (additional relationship)
Sole average age-0 size	13	NAOI	0.650	** (additional relationship)
Bass average age-0 size	13	NAOI	0.611	*
Smelt average age-0 size	11	NAOI	0.850	**
Smelt absolute growth	11	NAOI	0.838	**
Invertebrate measures				
ln(mean <i>Crangon</i> abundance)	16	NAOI	-0.636	**
ln(mean <i>Crangon</i> abundance)	16	LagNAOI	-0.574	*

Data were adjusted for trends (abundance data) or abundance (growth data) before statistical analysis. A single asterisk indicates significance at the 0.05 level; a double asterisk at the 0.01 level. Relationships additional to those presented in Table 1 are indicated. LagNAOI, NAOI with a lag period of 1 year. Plaice, *P. platessa*; sand goby, *P. minutus*.

contrast, several common fish species have lowest abundances during such years, and highest abundances during cold, low NAOI years. Analysis of water temperature data reveals that minimum and average winter temperatures are higher in the estuary than in the North Sea during low NAOI years, and vice versa (Fig. 2). Climatic conditions during years of low NAOI therefore result in estuarine areas becoming comparatively thermally favourable as nursery grounds, with higher numbers of fish using estuaries. This climate-driven pattern suggests a facultative rather than obligate use of the estuary by many juvenile fish²⁵, and highlights the importance of estuaries in providing recruits for certain fisheries (for example, herring, flatfish) during years of low NAOI. This mechanism is supported by the results for fish species that spend most of the year in estuaries (such as smelt, pipefish, poggie), and so would not be influenced by any temperature differential. The abundance of these species did not show any strong relationship with NAOI. De-trending of the data resulted in a greater number of significant population results than for regression alone (Tables 1 and 2). Significant correlations with de-trended series are thought to indicate a rapid response to the NAO⁷. The Thames data, therefore, provide strong evidence for the facultative use of estuarine environments in response to climate-induced temperature differentials, and a plausible mechanism through which the NAO can influence the status of marine fish populations.

Recruitment of fish from estuaries can strongly drive marine population dynamics²⁶, partly owing to their biological importance as nursery areas for many commercial fish species. Our results suggest that juvenile estuarine fish populations are strongly affected by climatic variability, as influenced by the NAO, in ways that may affect estuarine production potential through changes in either growth or abundance. In particular, the facultative use of estuaries implies that estuaries are important ecological buffers for many species in years of low NAOI. Therefore, estuaries are potentially critical for dampening climate-induced stock fluctuations. Knowledge of this role is important for sustainable management of fisheries, particularly during stock recovery²⁷, and highlights the importance of targeting conservation resources into estuaries to preserve the ecological buffering capacity of estuaries for commercially important or endangered estuarine-dependent fish stocks. □

Methods

Sampling methodology

Fish were sampled approximately every two weeks over a 16-year (1977–92) period from West Thurrock Power Station, north bank of the Thames estuary, UK²⁰. All species retained over a 4-h period at low spring tide were identified, counted, and individuals of commercially important species were measured¹⁷. We standardized abundance data to 500,000 m³ of filtered water²⁰. Data for estuarine environmental parameters (temperature, freshwater flow, salinity, dissolved oxygen, suspended solids, pH, nitrogen) coincident to biota samples were obtained from Environment Agency databases¹⁸. Southwest North Sea SST data were obtained from the German Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie). Winter NAOI data⁴ (the difference between the normalized mean winter (December–February) surface-pressure anomaly for Ponta Delgadas (Azores) and that for Akureyri (Iceland)) were provided by the Department of Fisheries and Oceans, Canada. To investigate annual trends, the yearly mean values for sample diversity indices, fish abundance and the abundance of individual species were calculated, together with the total number of fish caught per year.

Statistical analysis

Length–frequency data for individual sample dates were used to compute estuarine entry and exit sizes and absolute growth (cm) for the period of estuarine residency¹⁰. Length–frequency data for individual sample dates were used to compute estuarine entry and exit sizes and absolute growth (cm) for all juvenile fish less than one year old (age-0) for the period of estuarine residency. Mean values for estuarine environmental parameters (annual) and North Sea SST (monthly and annual) were also computed. Abundance data series were assessed for temporal trends before further use in statistical analysis using linear trend regressions. Insignificant slope coefficients (18 out of 22) or statistically inadequate regression models (4 out of 22) provided no compelling evidence of temporal trends in the data. Growth data were assessed for correlation with abundance before further use in statistical analysis using linear regression. We found no association between

growth and abundance. Biological data were regressed against NAOI values to assess the pervasiveness and significance of large-scale climate processes as influences on estuarine biota. Estimated regression model residuals were tested for statistical adequacy following standard practices including tests for normal, independent and homoscedastic residuals, all applied at the $\alpha = 0.05$ level of significance. Adjusted significance levels for each regression were computed using the Bonferroni inequality to test the joint significance of all NAOI regressions⁷ under the assumption that the joint probability of all results holding simultaneously was at least 0.90 ($\alpha = 0.10$), and where α was chosen to minimize corresponding type II error. Partial correlation coefficients for abundance and NAOI data that controlled for temporal trend (year), and growth and NAOI data that controlled for species abundance, were computed to verify regression results. We tested coefficients for significance using a two-tailed Student's *t*-test.

Received 5 July 2001; accepted 6 February 2002.

1. Corten, A. & van de Kamp, G. Variation in the abundance of southern fish species in the southern North Sea in relation to hydrography and wind. *ICES J. Mar. Sci.* **53**, 1113–1119 (1996).
2. Alheit, J. & Hagen, E. Long-term climate forcing of European herring and sardine populations. *Fish. Oceanogr.* **6**, 130–139 (1997).
3. O'Brien, C. M., Fox, C. J., Planque, B. & Casey, J. Climate variability and North Sea cod. *Nature* **404**, 142 (2000).
4. Rogers, J. C. The association between the North Atlantic Oscillation and the Southern Oscillation in the northern hemisphere. *Mon. Weath. Rev.* **112**, 1999–2015 (1984).
5. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperature and precipitation. *Science* **269**, 676–679 (1995).
6. Reid, P. C., Planque, B. & Edwards, M. Is observed variability in the long-term results of the Continuous Plankton Recorder survey a response to climate change? *Fish. Oceanogr.* **7**, 282–288 (1998).
7. Ottersen, G. *et al.* Ecological effects of the North Atlantic Oscillation. *Oecologia* **128**, 1–14 (2001).
8. Dippner, J. W. Recruitment success of different fish stocks in the North Sea in relation to climate variability. *Ger. J. Hydrogr.* **49**, 277–293 (1997).
9. Hovgård, H. & Buch, E. In *Large Marine Ecosystems: Patterns, Processes and Yields* (eds Sherman, K., Alexander, L. M. & Gold, B. D.) 36–43 (American Association for the Advancement of Science (AAAS), Washington, 1992).
10. Wootton, R. J. *Ecology of Teleost Fishes* (Chapman & Hall, London, 1990).
11. Blaber, S. J. M. & Blaber, T. G. Factors affecting the distribution of juvenile estuarine and inshore fishes. *J. Fish Biol.* **17**, 143–162 (1980).
12. Haddon, M. *Modelling and Quantitative Methods in Fisheries* (Chapman & Hall/CRC, Boca Raton, Florida, 2001).
13. Hilborn, R. & Walters, C. J. *Quantitative Fisheries Stock Assessment* (Chapman & Hall, London, 1992).
14. Brander, K. M. The effects of temperature on growth of Atlantic cod (*Gadus morhua* L.). *ICES J. Mar. Sci.* **52**, 1–10 (1995).
15. Drinkwater, K. F. & Myers, R. A. Testing predictions of marine fish and shellfish landings from environmental variables. *Can. J. Fish. Aquat. Sci.* **44**, 1568–1573 (1987).
16. Rodwell, M. J., Rodwell, D. P. & Folland, C. K. Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature* **398**, 320–323 (1999).
17. Power, M., Attrill, M. J. & Thomas, R. M. Temporal abundance patterns and growth of juvenile herring and sprat from the Thames estuary 1977–1992. *J. Fish Biol.* **56**, 1408–1426 (2000).
18. Attrill, M. J. & Power, M. Effects on invertebrate populations of drought-induced changes in estuarine water quality. *Mar. Ecol. Prog. Ser.* **203**, 133–143 (2000).
19. Power, M., Attrill, M. J. & Thomas, R. M. Environmental factors and interactions affecting the temporal abundance of juvenile flatfish in the Thames Estuary. *J. Sea Res.* **43**, 135–149 (2000).
20. Thomas, R. M. In *A Rehabilitated Estuarine Ecosystem: the Environment and Ecology of the Thames Estuary* (ed. Attrill, M. J.) 115–140 (Kluwer, Dordrecht, 1998).
21. Kröncke, I., Dippner, J. W., Heyen, H. & Zeiss, B. Long-term changes in macrofaunal communities off Nordeney (East Frisia, Germany) in relation to climate variability. *Mar. Ecol. Prog. Ser.* **167**, 25–36 (1998).
22. Blaxter, J. H. S. & Hunter, J. R. The biology of dupeoid fishes. *Adv. Mar. Biol.* **20**, 1–224 (1982).
23. Wheeler, A. *The Fishes of the British Isles and North West Europe* (Michigan State Univ. Press; Macmillan, London, 1969).
24. Bettoso, N. & Dulcic, J. First record of the oilfish *Ruvettus pretiosus* in the northern Adriatic Sea. *J. Mar. Biol. Ass. UK* **79**, 1145–1146 (1999).
25. Potter, I. C., Gardner, D. C. & Claridge, P. N. Age composition, growth, movements, meristics and parasites of the whiting *Merlangius merlangus*, in the Severn Estuary and Bristol Channel. *J. Mar. Biol. Ass. UK* **68**, 295–313 (1988).
26. Elliott, M. & Taylor, C. J. L. In *Proceedings of the 21st European Marine Biology Symposium, Gdansk, 14–19 September 1986* (eds Klekowski, R. Z., Styczynska-Jurewicz, E. & Falkowski, L.) 227–240 (Olsen and Olsen, Fredensborg, 1989).
27. Hutchings, J. A. Collapse and recovery of marine fisheries. *Nature* **406**, 882–885 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank the Environment Agency for permission to analyse and publish the data. Support for completion of the work was provided by an NSERC research grant to M.P.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.A. (e-mail: matrill@plymouth.ac.uk).

Nonlinear grassland responses to past and future atmospheric CO₂

Richard A. Gill^{*†}, H. Wayne Polley[‡], Hyrum B. Johnson[‡],
Laurel J. Anderson^{†§}, Hafiz Maherali^{*} & Robert B. Jackson^{*||}

^{*} Department of Biology; and ^{||} Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708-0340 North Carolina, USA

[‡] USDA-ARS Grassland, Soil and Water Research Laboratory, Temple, Texas 76502-9601, USA

[§] Department of Botany, University of Texas, Austin, Texas 78713, USA

[†] Present addresses: Program in Environmental Science and Regional Planning, Washington State University, Pullman, Washington 99164, USA (R.A.G.); Ohio Wesleyan University, Department of Botany-Microbiology, Delaware, Ohio 43015, USA (L.J.A.)

Carbon sequestration in soil organic matter may moderate increases in atmospheric CO₂ concentrations (C_a) as C_a increases to more than 500 μmol mol⁻¹ this century from interglacial levels of less than 200 μmol mol⁻¹ (refs 1–6). However, such carbon storage depends on feedbacks between plant responses to C_a and nutrient availability^{7,8}. Here we present evidence that soil carbon storage and nitrogen cycling in a grassland ecosystem are much more responsive to increases in past C_a than to those forecast for the coming century. Along a continuous gradient of 200 to 550 μmol mol⁻¹ (refs 9, 10), increased C_a promoted higher photosynthetic rates and altered plant tissue chemistry. Soil carbon was lost at subambient C_a, but was unchanged at elevated C_a where losses of old soil carbon offset increases in new carbon. Along the experimental gradient in C_a there was a nonlinear, threefold decrease in nitrogen availability. The differences in sensitivity of carbon storage to historical and future C_a and increased nutrient limitation suggest that the passive sequestration of carbon in soils may have been important historically, but the ability of soils to continue as sinks is limited.

The concentration of CO₂ in the atmosphere has increased dramatically since the Last Glacial Maximum, most recently owing to fossil fuel burning and land conversion to agriculture.

This increase in C_a has focused attention on the role of terrestrial ecosystems in sequestering anthropogenic CO₂ (refs 2, 5, 7, 11, 12). The long-term consequences of rising C_a on C sequestration are highly dependent on feedbacks between plant responses to C_a and nutrient dynamics^{7,8,13}. Plant growth is often enhanced with increases in C_a (refs 6, 14), sometimes leading to changes in plant tissue chemistry and organic inputs to soils^{15,16}. These and other feedbacks controlled by microbial processes may either increase^{13,17} or decrease^{7,8,18} nutrient availability, and mediate the long-term ability of ecosystems to sequester C^{7,8,19}. For C sequestration to be important at decadal and century timescales, nutrient availability must not hinder higher plant production and new organic C must be stabilized in soil pools with relatively long turnover times. The partitioning of C among soil organic matter (SOM) pools with different turnover rates is thus a crucial determinant of C sequestration in many systems and is tightly coupled with plant tissue chemistry and nutrient dynamics^{13,16,18}.

A field experiment⁹ in an intact C₃/C₄ grassland in central Texas provided a continuous gradient of C_a from 200 to 550 μmol mol⁻¹ permitting the measurement of critical threshold and nonlinear responses to past, present and future atmospheric CO₂. Plant and ecosystem properties, including water-use efficiency, photosynthesis, respiration rates and primary productivity, often change with rising C_a, but it is not likely that all such responses were or will be linear^{3,20,21}. Physiological thresholds²⁰, transient or acclimatory responses²², and the strong coupling of plant and soil responses¹⁸ are examples of mechanisms that may drive nonlinear processes in nature²³. Nonlinear and threshold responses are the focus of several new international programmes²³ and may explain some of the apparent contradictory results observed in recent CO₂ studies^{8,13}. Furthermore, research on how intact ecosystems respond to both past and future C_a provides a context that can demonstrate the sensitivity of C dynamics to changes that have already occurred as well as those forecast for the coming century. Extrapolation from experiments that impose step changes in C_a is complicated by the possibility that plants may evolve as C_a changes more slowly in nature. There is some evidence, however, that perennial plants have not evolved quickly enough to be closely adapted to current C_a (ref. 24).

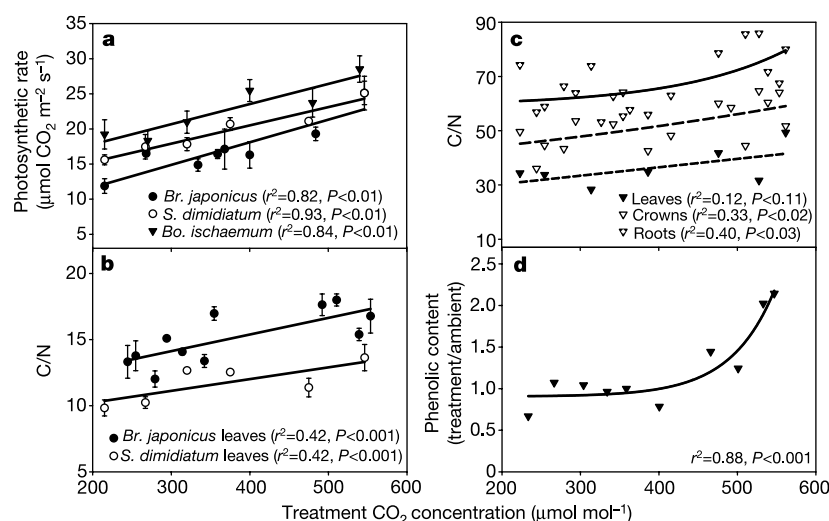


Figure 1 Effects of CO₂ treatment on various species. **a**, Maximum CO₂ assimilation for three species (*Bothriochloa ischaemum*, *Solanum dimidiatum*, *Bromus japonicus*) in 1999, showing a significant positive relationship between maximum CO₂ assimilation and treatment CO₂ in all species ($P < 0.01$). **b**, C/N ratio for leaves from the two C₃ species show a positive, linear increase for both species with increasing treatment CO₂. **c**, C/N ratio for *Bo. ischaemum* roots, crowns and leaves. Roots showed an exponential increase

in C/N ratio with increasing CO₂ ($P < 0.03$); crowns showed a positive, linear increase ($P < 0.05$). **d**, Relative change in phenolic concentrations in *Bo. ischaemum* roots (expressed relative to ambient values). There was a strong, exponential increase in root phenolic content ($P < 0.001$), with an apparent threshold at C_a slightly above ambient levels.

Table 1 Calculated carbon production and storage

	Year	Elevated (550–350 $\mu\text{mol mol}^{-1}$)	Subambient (365–200 $\mu\text{mol mol}^{-1}$)	R^2 (P value)	Superambient/ subambient	Pretreatment superambient/subambient
Aboveground net primary production ($\text{g m}^{-2} \text{yr}^{-1}$)	1996–2000	1,047.5 (64.9)	683.9 (52.2)	0.35 (0.006)	1.52	0.83
Belowground net primary production 0–30 cm ($\text{g m}^{-2} \text{yr}^{-1}$)	1998–1999	294 (24.6)	185 (22.8)	0.16 (0.09)	1.59	1.03
Soil CO_2 flux ($\mu\text{g m}^{-2} \text{sec}^{-1}$)	1996–2000	4.02 (0.13)	2.85 (0.17)	0.46 (0.001)	1.41	1.13
Root biomass 0–30 cm (g m^{-2})	1998	173.0 (39.4)	102.0 (16.2)	0.25 (0.02)	1.69	0.84
Soil organic carbon 0–15 cm (g m^{-2})	1996–2000	4,442 (175)	3,656 (120)	0.32 (0.05)	1.22	1.05

R^2 and P values are from best-fit regressions of variables on C_a over subambient through elevated concentrations. s.e.m., the standard error of the mean, shown in parentheses after the mean value. For all data other than root biomass, s.e.m. is determined as the standard error for annual means; for root biomass it is the standard error between section means ($n = 10$).

Along the experimental gradient, plants responded to higher C_a by increasing photosynthesis and net primary production (Fig. 1a, Table 1). As treatment CO_2 increased, maximum CO_2 assimilation rates increased linearly for both C_3 and C_4 plants¹⁰ (Fig. 1a; $P < 0.01$). Associated with this increase in CO_2 assimilation was a 50% increase in above- and belowground net primary production at elevated CO_2 compared to subambient CO_2 (Table 1). Tissue chemistry was altered as well, with an increase in tissue C/N with higher C_a and an exponential increase in phenolic concentration (Fig. 1b–d). C_a and species type were highly significant predictors of C/N, with C/N positively correlated with C_a (analysis of covariance (ANCOVA); $P < 0.001$ for C_a ; $P < 0.001$ for species). The concentration of phenolic compounds in roots of one of the dominant species in the system, the C_4 grass *Bothriochloa ischaemum*, showed a strong threshold effect, with little variation in plants grown at subambient C_a , but an exponential increase above ambient CO_2 (Fig. 1d, $P < 0.001$).

Soil C storage and belowground metabolism were greatly altered. Despite a linear increase in photosynthesis along the gradient, soil C storage was much more sensitive to subambient than to elevated C_a (Fig. 2a). At subambient C_a , bulk soil C stocks decreased by 11%, or 450 g m^{-2} , between 1996 and 2000 (Table 2). However, there was no concomitant increase in soil C storage at elevated C_a (Fig. 2a), with soil C increasing by a modest 3.3% (144 g m^{-2}) over the same time period (Table 2). The relationship between treatment CO_2 and the change in bulk soil organic C over three years follows an asymptotic function (Fig. 2a, $P < 0.05$), suggesting that the ability of soils to act as sinks for anthropogenic CO_2 will slow or reach saturation.

Accompanying altered soil C storage was an important change in soil organic matter chemistry. Total organic matter C/N was linearly associated with treatment C_a (Fig. 2c, $P < 0.01$), in a pattern similar to that observed for plant tissue chemistry. There was also a divergence in patterns of soil respiration at super- versus subambient C_a . Soil CO_2 flux at peak plant growth was 40% higher at elevated than at subambient C_a , suggesting that much of the increase in C fixed with rising C_a is lost to microbial or root respiration⁵ (Table 1).

The changes observed in particulate organic matter (POM) demonstrate a shift in the balance between new and old SOM. POM is a relatively labile class of SOM, with a residence time of between 10 and 50 years^{11,25,26}. The 14% loss in POM carbon at subambient C_a parallels the loss in total organic C (Table 2). However, in contrast to total organic C, POM C increased linearly with treatment CO_2 , even at elevated C_a (Fig. 2b). These findings indicate that at elevated C_a , increases in POM C were largely offset by losses in the older, mineral-associated organic matter²⁷ (Table 2). Even within the POM class, there were increases at elevated C_a in the two most labile fractions (free and macroaggregate POM), while there was a decrease in the most recalcitrant fraction (microaggregate POM)²⁶ (Table 2). This represents a change in ecosystem C partitioning to faster cycling organic matter^{11,16,26}, which may explain why higher C assimilation and production did not lead to increased C sequestration. Our result is similar to those of other studies that reported that at low nutrient availability and elevated CO_2 , carbon was lost from the mineral-bound fraction of SOM²⁵. Similarly, an annual grassland exposed to a doubling of C_a had

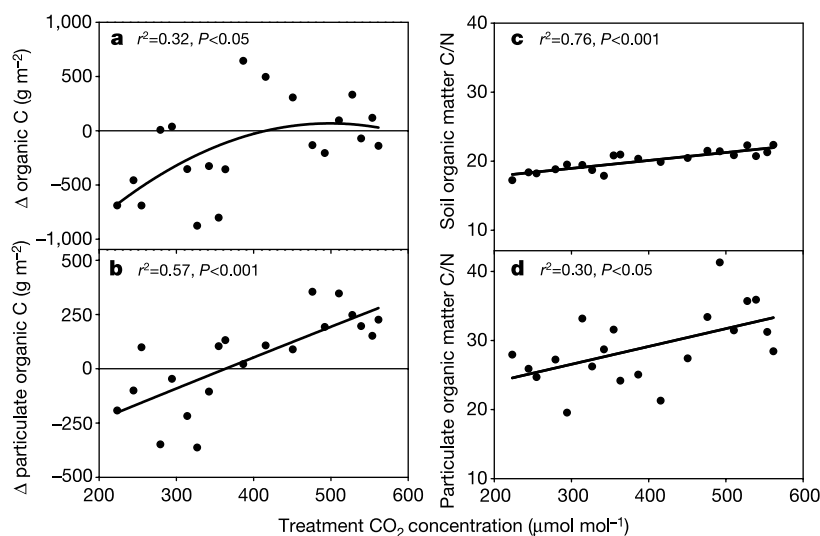


Figure 2 Effect of CO_2 treatment on soil carbon storage. **a**, Change in organic C stocks (0–15 cm) between 1997 and 2000. Values are the difference between section means in 1997 and 2000 determined using four subsamples per 5-m section per year. There was a quadratic relationship between the change in soil C stocks and treatment CO_2 ($P < 0.05$). The linear fit for these data was not significant. **b**, Significant, linear change in particulate

organic matter (POM) carbon between 1997 and 2000 ($P < 0.001$). **c**, There was a significant, linear increase in bulk soil organic matter (SOM) C/N with treatment CO_2 in December 2000 ($P < 0.001$). **d**, POM C/N for samples collected in December 2000 ($P < 0.05$). Values are the means of four subsamples per section.

Table 2 Pools and changes in soil organic carbon and particulate organic carbon

	Treatment leg	December 2000 0–15 cm (g m ⁻²)	Change (g m ⁻²) 1997–2000	Relative change (%)
Total soil organic matter C	Superambient	4,442 (175)	144 (92)	3.3
	Subambient	3,656 (120)	-450 (100)	-11
Particulate organic matter C—Free	Superambient	186.1 (23.7)	86.5 (22.0)	70
	Subambient	158.2 (16.8)	-26.9 (6.3)	-16
Macroaggregate	Superambient	723.4 (38.7)	193.3 (33.9)	36
	Subambient	626.6 (39.8)	-104.0 (56.8)	-14
Microaggregates	Superambient	87.5 (6.7)	-21.6 (22.6)	-24
	Subambient	65.9 (6.4)	-5.3 (5.5)	-9
Total particulate organic matter C	Superambient	975.4 (33.9)	258.2 (70.1)	35
	Subambient	857.3 (56.8)	-132.3 (68.2)	-14
Mineral-associated organic matter C	Superambient	3,719 (195)	-123 (96)	-3.3
	Subambient	3,030 (118)	-346 (121)	-9.8

s.e.m., the standard error of the section means ($n = 10$), shown in parentheses after the mean value.

higher ecosystem C uptake and belowground allocation but little extra C storage⁵. Much of the increased C was partitioned to rapidly cycling pools that make a negligible contribution to long-term storage because of their small size and relatively high turnover rates.

The feedback between plant responses to C_a and nutrient dynamics is vital in determining C sequestration in ecosystems^{7,8,18}. Nitrogen mineralization rates decreased dramatically and non-linearly with increasing CO_2 ($P < 0.01$), with the largest changes occurring at subambient concentrations (Fig. 3). Net N mineralization was three times higher at 200–240 $\mu\text{mol mol}^{-1}$ CO_2 than at 530–550 $\mu\text{mol mol}^{-1}$. Because of the changes in the chemical composition of detritus and increased C supply, microbes at high CO_2 may need to mineralize older, mineral-associated SOM to meet their nutritional requirements. As a result, there was a decrease in plant-available N as a consequence of microbial immobilization and a loss in C stored in mineral-associated fractions of organic matter. Some workers have concluded that suppressed N availability under elevated CO_2 may increase C storage by suppressing decomposition rates^{8,18}, but we found that there were only modest gains in soil C storage at the lowest N availability. In contrast to other grassland

CO_2 studies^{6,14,21}, our results are apparently a consequence of altered plant litter chemistry rather than an indirect effect of altered soil water status, as increases in plant water-use efficiency along the gradient¹⁰ were offset by higher plant biomass (data not shown). Increases in C_a resulted in higher nitrogen-use efficiency by plants¹⁰, but a threefold decrease in nitrogen availability will probably have a detrimental effect on long-term plant productivity and, ultimately, on ecosystem carbon storage.

Higher net primary productivity^{5,7}, altered plant tissue chemistry²⁷, modifications of SOM composition and stocks^{5,11,25}, and changes in nutrient availability^{13,18} with increases in C_a suggest that both forests and grasslands are sensitive to rising CO_2 . The capacity of future ecosystems to act as sinks for anthropogenic CO_2 will be determined by feedbacks among ecosystem processes^{7,18} and will be sensitive to the location of specific thresholds that influence the magnitude of the change in ecosystem dynamics²³. In this grassland, soil C stocks and net N mineralization are much more sensitive to subambient than elevated C_a , indicating that we are currently at an important threshold. Soils may have played a role in passively sequestering C since the last interglacial period, but their ability to continue to act as a C sink may be limited by nutrient availability. To assess the impacts of rising CO_2 on carbon sequestration patterns and nutrient dynamics requires knowledge of potential threshold responses and the legacy of historical and prehistorical changes. □

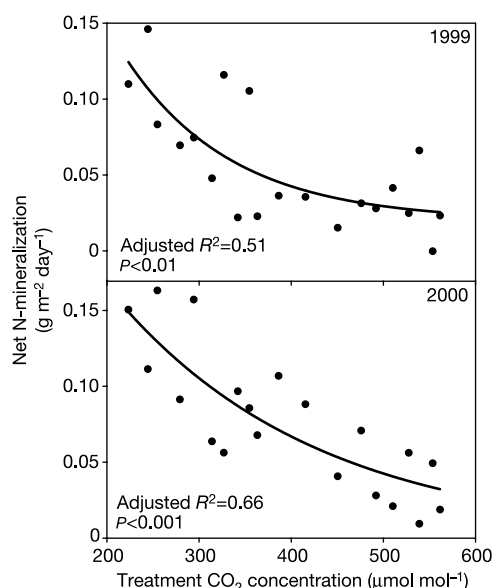


Figure 3 Net N mineralization during the 1999 and 2000 growing seasons. There was a significant, negative exponential relationship between net N mineralization and treatment CO_2 during midsummer in both years ($P < 0.001$). During spring and autumn there were no significant differences in N mineralization rates for the subambient and elevated chambers.

Methods

Experimental system

Two parallel, elongated chambers (1 m tall $\times$ 1 m wide $\times$ 60 m long) were constructed on a grassland dominated by the C_4 perennial grass *Bothriochloa ischaemum* (L.) Keng and Ambient air plus the C_3 perennial forbs *Solanum dimidiatum* Raf. and *Ratibida columnaris* (Sims) D. Don. pure CO_2 was injected into the eastern chamber to initiate the elevated gradient (550–350 $\mu\text{mol mol}^{-1}$), while ambient air was injected into the western chamber, initiating the subambient CO_2 gradient (365–200 $\mu\text{mol mol}^{-1}$). Gradients have been maintained during the growing season since May 1997 by altering flow rate through the chambers. At night, air flow in the chambers is reversed, maintaining a C_a gradient at 150 $\mu\text{mol mol}^{-1}$ above daytime concentrations. The chambers are divided into 5-m sections, and air is cooled and dehumidified in each section to maintain air temperature and vapour pressure deficit near ambient conditions. Our results span pre-treatment data (1996–1997) and the three complete growing seasons during which the grassland was exposed to a C_a gradient (1998–2000).

Soil analyses

Soil respiration was evaluated monthly using a LI-COR 6200. Total inorganic and organic soil carbon was determined using a two-temperature combustion procedure designed specifically for calcareous Blackland Prairie soils²⁸. Four soil cores were collected from each of the 20 sections in stratified, random positions. Total C and N were measured using a CE Instruments NC 2100 elemental analyser (ThermoQuest Italia). We measured POM in two aggregate size classes (macroaggregates ($>250 \mu\text{m}$); microaggregates (250–53 μm)) using the method described in ref. 26 to determine POM C. Mineral-associated C was determined by difference between total C and POM C. We determined POM C using four soil samples from each section ($n = 80$) that were collected in September 1997 and December 2000. We used a month-long, *in situ* open-core incubation method described in ref. 29 to measure net nitrogen mineralization.

Statistical considerations

The experimental system is constructed to resolve the shape of ecosystem responses to a gradient in CO₂. The experimental design uses a regression approach to test for significant CO₂ effects based on changes in slope along the gradient. We used regression to test for a significant relationship between C_a and the response variable using the regression wizard in SigmaPlot 5.0 for Windows (SPSS Inc.). We tested linear, logarithmic, power and hyperbolic functions to fit the data, and selected the model with the highest adjusted R² after examining the residual plots for normality and homoscedasticity. When models were nearly the same in their explanatory value (R² values within 0.05), we report results for the linear model. Because we had only a single experimental system oriented in one direction across the landscape, it is possible that the measured responses may have been influenced by some unquantified factor covarying with CO₂ treatment. However, extensive pretreatment data, including such ecosystem characteristics as soil C stocks, net primary productivity and soil respiration, revealed no such trends before fumigation (Table 1 and additional data not shown). Furthermore, the system design ensured that key environmental variables (photosynthetically active radiation, T, relative humidity, and so on) remained similar across the gradient⁹. The absence of strong threshold responses at the transition between the two chambers provides further evidence that neither landscape position nor position within the chamber significantly influenced observations. To control for any pre-existing variation in soil organic matter, we evaluate the change in soil C stocks between 1997 and 2000 rather than absolute levels (Table 2).

Received 19 July 2001; accepted 14 February 2002.

- Keeling, C. D., Chini, J. F. S. & Whorf, T. P. Increased activity of northern vegetation inferred from atmospheric CO₂ measurements. *Nature* **382**, 146–149 (1996).
- Schimmel, D. S. Terrestrial ecosystems and the carbon cycle. *Glob. Change Biol.* **1**, 77–91 (1995).
- Polley, H. W., Johnson, H. B., Marino, B. D. & Mayeux, H. S. Increase in C3 plant water-use efficiency and biomass over glacial to present CO₂ concentrations. *Nature* **361**, 61–64 (1993).
- Follett, R. F., Kimble, J. M. & Lal, R. In *The Potential of U.S. Grazing Lands to Sequester Carbon and Mitigate the Greenhouse Effect* (eds Follett, R. F., Kimble, J. M. & Lal, R.) 401–430 (Lewis, Boca Raton, 2001).
- Hungate, B. A. *et al.* The fate of carbon in grasslands under carbon dioxide enrichment. *Nature* **388**, 576–579 (1997).
- Owensby, C. E., Ham, J. M., Knapp, A. K. & Auen, L. M. Biomass production and species composition change in a tallgrass prairie ecosystem after long-term exposure to elevated atmospheric CO₂. *Glob. Change Biol.* **5**, 497–506 (1999).
- Oren, R. *et al.* Soil fertility limits carbon sequestration by a forest ecosystem in a CO₂-enriched atmosphere. *Nature* **411**, 469–472 (2001).
- Diaz, S., Grime, J. P., Harris, J. & McPherson, E. Evidence of feedback mechanism limiting plant response to elevated carbon dioxide. *Nature* **364**, 616–617 (1993).
- Johnson, H. B., Polley, H. W. & Whitis, R. P. Elongated chambers for field studies across atmospheric CO₂ gradients. *Funct. Ecol.* **14**, 388–396 (2000).
- Anderson, L. J., Maherali, H., Johnson, H. B., Polley, H. W. & Jackson, R. B. Gas exchange and photosynthetic acclimation over subambient to elevated CO₂ in a C₃-C₄ grassland. *Glob. Change Biol.* **7**, 693–707 (2002).
- Schlesinger, W. H. & Lichter, J. Limited carbon storage in soils and litter of experimental forest plots under increased atmospheric CO₂. *Nature* **411**, 466–469 (2001).
- Tans, P. P. & White, J. W. C. In balance, with a little help from the plants. *Science* **281**, 183–184 (1998).
- Zak, D. R. *et al.* Elevated atmospheric CO₂ and feedback between carbon and nitrogen cycles. *Plant Soil* **151**, 105–117 (1993).
- Morgan, J. A., LeCain, D. R., Mosier, A. R. & Milchunas, D. G. Elevated CO₂ enhances water relations and productivity and affects gas exchange in C₃ and C₄ grasses of the Colorado shortgrass steppe. *Glob. Change Biol.* **7**, 451–466 (2001).
- Canadell, J. G., Pitelka, L. F. & Ingram, J. S. The effects of elevated CO₂ on plant-soil carbon belowground. *Plant Soil* **187**, 391–400 (1996).
- Van Kessel, C. *et al.* Carbon-13 input and turn-over in a pasture soil exposed to long-term elevated atmospheric CO₂. *Glob. Change Biol.* **6**, 123–135 (2000).
- Körner, Ch. & Arnone, J. A. Response to elevated carbon dioxide in artificial tropical ecosystems. *Science* **257**, 1672–1675 (1992).
- Hu, S., Chapin, F. S., Firestone, M. K., Field, C. B. & Chiariello, N. R. Nitrogen limitation of microbial decomposition in a grassland under elevated CO₂. *Nature* **409**, 188–191 (2001).
- Zak, D. R. *et al.* Atmospheric CO₂, soil-N availability, and allocation of biomass and nitrogen by *Populus tremuloides*. *Ecol. Appl.* **10**, 34–46 (2000).
- Hättenschwiler, S. & Körner, C. Effects of elevated CO₂ and increased nitrogen deposition on photosynthesis and growth of understory plants in spruce model ecosystems. *Oecologia* **106**, 172–180 (1996).
- Jackson, R. B., Sala, O. E., Field, C. B. & Mooney, H. A. CO₂ alters water use, carbon gain, and yield for the dominant species in a natural grassland. *Oecologia* **98**, 257–262 (1994).
- Oechel, W. C. *et al.* Transient nature of CO₂ fertilization in Arctic tundra. *Nature* **371**, 500–503 (1994).
- Körner, C. Biosphere responses to CO₂ enrichment. *Ecol. Appl.* **10**, 1590–1619 (2000) (see Global Change and Terrestrial Ecosystems at <http://www.gcte.org>; and International Geosphere-Biosphere Programme at <http://www.igbp.kva.se>).
- Sage, R. E. & Cowling, S. A. In *Carbon Dioxide and Environmental Stress* (eds Luo, Y. & Mooney, H. A.) 289–308 (Academic, San Diego, 1999).
- Cardon, Z. G. *et al.* Contrasting effects of elevated CO₂ on old and new soil carbon pools. *Soil Biol. Biochem.* **33**, 365–373 (2001).
- Six, J., Elliott, E. T., Paustian, K. & Doran, J. W. Aggregation and soil organic matter accumulation in cultivated and native grassland soils. *Soil Sci. Soc. Am. J.* **62**, 1367–1377 (1998).
- Norby, R. J., Cotrufo, M. F., Ineson, P., O'Neill, E. G. & Canadell, J. G. Elevated CO₂, litter chemistry, and decomposition: a synthesis. *Oecologia* **127**, 153–165 (2001).
- Chichester, F. W. & Chaison, R. F. Jr Analysis of carbon in calcareous soils using a two temperature dry combustion infrared instrumental procedure. *Soil Sci.* **153**, 237–241 (1992).
- Hook, P. B. & Burke, I. C. Evaluation of methods for estimating net nitrogen mineralization in a semiarid grassland. *Soil Sci. Soc. Am. J.* **59**, 831–837 (1995).

Acknowledgements

We thank W. Pockman, W. Gordon and S. Brumbaugh for assistance in the field; R. Cates for liquid chromatography analysis; R. P. Whitis, C. W. Cook and A. Gibson for technical assistance; and W. K. Schlesinger, B. Hungate, E. Jobbagy, J. Powers and A. Finzi for comments on the manuscript. This paper is a contribution to the Global Change and Terrestrial Ecosystems core project of the International Geosphere Biosphere Programme. This research was supported by the National Institute for Global Environmental Change through the US Department of Energy (R.B.J.) and the US Department of Agriculture National Research Initiative Competitive Grants Program (R.A.G.). Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the DOE, National Institute for Global Environmental Change or the National Research Initiative Competitive Grants Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.G. (e-mail: rgill@wsu.edu).

Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning

J. Amiel Rosenkranz* & Anthony A. Grace*†

* Department of Neuroscience and † Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

Pavlovian conditioning results when an innocuous stimulus, such as an odour, is paired with a behaviourally relevant stimulus, such as a foot-shock, so that eventually the former stimulus alone will elicit the behavioural response of the latter. The lateral nucleus of the amygdala (LAT) is necessary for the emotional memory formation in this paradigm^{1–4}. Enhanced neuronal firing in LAT to conditioned stimuli emerge in parallel with the behavioural changes^{5–11} and are dependent on local dopamine^{12–15}. To study the changes in neuronal excitability and synaptic drive that contribute to the pavlovian conditioning process, here we used *in vivo* intracellular recordings to examine LAT neurons during pavlovian conditioning in rats. We found that repeated pairings of an odour with a foot-shock resulted in enhanced post-synaptic potential (PSP) responses to the odour and increased neuronal excitability. However, a non-paired odour displayed PSP decrement. The dopamine antagonist haloperidol blocked the PSP enhancement and associated increased neuronal excitability, without reversing previous conditioning. These results demonstrate that conditioning and habituation processes produce opposite effects on LAT neurons and that dopamine is important in these events, consistent with its role in emotional memory formation.

In male rats (Sprague–Dawley, 250–350 g) anaesthetized with 8% chloral hydrate, odour-evoked depolarizing responses can be observed in neurons of the LAT, as well as responses evoked by a foot-shock (Fig. 1). Repeated presentation of an odour resulted in a gradual attenuation of the odour-evoked PSPs (Fig. 2; $n = 4$, $P < 0.01$, $F = 10.9$, degrees of freedom (d.f.) = 6, repeated measures analysis of variance (ANOVA)), as well as causing a significant suppression of membrane fluctuations during the odour presentation to below the level of spontaneous activity (baseline odour-evoked PSP $3,287.5 \pm 459.2$ mV ms; after six presentations $-1,302.3 \pm 175.7$ mV ms). In contrast, in a separate group, pairing of an odour with a train of foot-shocks in a pavlovian

conditioning procedure resulted in an increase in the response to the paired odour ($n = 25$, $P < 0.01$, $F = 4.99$, d.f. = 7, repeated measures ANOVA, Fig. 2). Alternatively, in cases in which the response to the odour was initially undetected, pairing of the odour with the foot-shock caused the response to the odour to emerge. The mean enhancement of odour-evoked responses was approximately 400% (baseline odour-evoked PSP $1,062.1 \pm 590.9$ mV ms; post-conditioning $4,331.6 \pm 599.5$ mV ms; $P < 0.01$, $t = -577$, d.f. = 24, paired t -test). The specificity of conditioning to the paired odour was verified by the absence of enhanced response to a second, non-paired odour (baseline response $1,216.0 \pm 559.5$ mV ms; post-conditioning 230.3 ± 250.6 mV ms; $P = 0.11$, $t = 1.65$, d.f. = 24, paired t -test), thereby controlling for nonspecific effects of the conditioning procedure¹⁶. Before conditioning there was no significant difference in the response to the paired and non-paired odours ($P = 0.85$, $t = -0.19$, d.f. = 48, t -test); however, after conditioning, the paired odour evoked a greater-magnitude PSP than the non-paired odour ($P < 0.01$, $t = 6.31$, d.f. = 48, t -test, Fig. 2).

These data demonstrate that *in vivo* electrophysiological correlates of affective pavlovian conditioning can be observed in anaesthetized rats. The contrasting effects of the habituation and conditioning procedures indicate that LAT neurons display correlates of both of these opponent processes. Indeed, the habituation procedure itself may assign a significance to the odour; that is, that

this odour is not contingent with any aversive event, and should not lead to an affective response. The result is an active suppression of LAT drive evoked by this odour, which would thereby minimize its influence upon affective behaviours. The enhanced post-synaptic response to the conditioned odour would be expected to drive LAT neurons, and thereby drive affective behaviour. The habituation of responses to non-paired odours, when presented in a fashion identical to the conditioning protocol, demonstrates that the enhancement is not due entirely to repeated exposure. The odour-evoked depolarizing response is probably mediated by a fast excitatory neurotransmitter, as LAT neurons often displayed a rapid response to the odour (<500 ms), and the odour-evoked depolarization resulted in action potential generation (in 7/25 cases).

To address whether membrane depolarization of neurons in the LAT is an integral component of this plasticity, the conditioning procedure was performed with the addition of 5-s-duration hyperpolarizing intracellular pulses (-0.2 to -0.43 nA, mean -0.27 ± 0.1 nA) coincident with the foot-shock. Hyperpolarizing pulses (mean -11.5 ± 1.7 mV) were adjusted to compensate for foot-shock-evoked depolarizations such that the neuronal membrane potential during the foot-shock remained close to, or hyperpolarized relative to the resting membrane potential (Fig. 2e). Addition of hyperpolarizing pulses significantly attenuated the enhancement of paired odours (control 25/31 neurons, hyperpolarization 0/5 neurons, control baseline response to odour $1,528.1 \pm 355.8$ mV ms; post-conditioning + hyperpolarization 262.5 ± 531.6 mV ms; $P < 0.05$, $t = 2.47$, paired t -test, d.f. = 9; Fig. 2f), but had no effect on non-paired odours (baseline response to odour $1,543.2 \pm 173.2$ mV ms; post-conditioning $1,364.1 \pm 275.0$ mV ms). However, neurons that displayed an initial odour-evoked response did not display attenuation to the degree observed after habituation (conditioning with hyperpolarizing current -82% , mean decrease $1,265.6$ mV ms; habituation -139% , mean decrease $4,589.8$ mV ms), implying that hyperpolarizing current reduced, but did not completely block conditioning, perhaps owing to a lesser effect of somatic current injection on dendritic compartments. These data indicate that post-synaptic somatic/proximal dendritic depolarization is a necessary component in these studies and, furthermore, that a portion of the enhancement of synaptic drive is due to changes in LAT neurons.

Previous studies have had variable success in demonstrating behavioural expression of affective conditioning procedures that were performed while the animal was anaesthetized^{17–20}. Two potential causes exist: lack of conditioning or lack of consolidation. Our results indicate that lack of consolidation is more likely, as plasticity correlated with the initial conditioning does occur. Consolidation of affective conditioning requires interactions between cortex and the amygdala²¹, which may be disrupted by anaesthesia. These data also indicate that peripheral autonomic responses may not be necessary for the initial formation of amygdala-related plasticity, as the anaesthesia levels utilized in this study block foot-shock-induced alterations of autonomic responses.

In a separate series of experiments the effects of dopamine receptor antagonism on this conditioned enhancement of the odour-evoked PSP was examined. In these experiments one odour was paired with foot-shock, followed by saline or haloperidol administration, and then the second odour was paired with the foot-shock. As above, pairing of the first odour with foot-shock resulted in an enhancement of the response to this odour (Fig. 3 and Table 1). When saline (0.4 ml of 0.9%, $n = 4$) was administered and the conditioning procedure was repeated with the second odour, a similar enhancement of the odour-evoked PSP was noted (Fig. 3 and Table 1). However, administration of haloperidol (0.67 ± 0.02 mg kg⁻¹ intravenously, $n = 7$) completely blocked the conditioned enhancement to the second odour (Fig. 3 and Table 1) without significantly affecting the previous enhanced

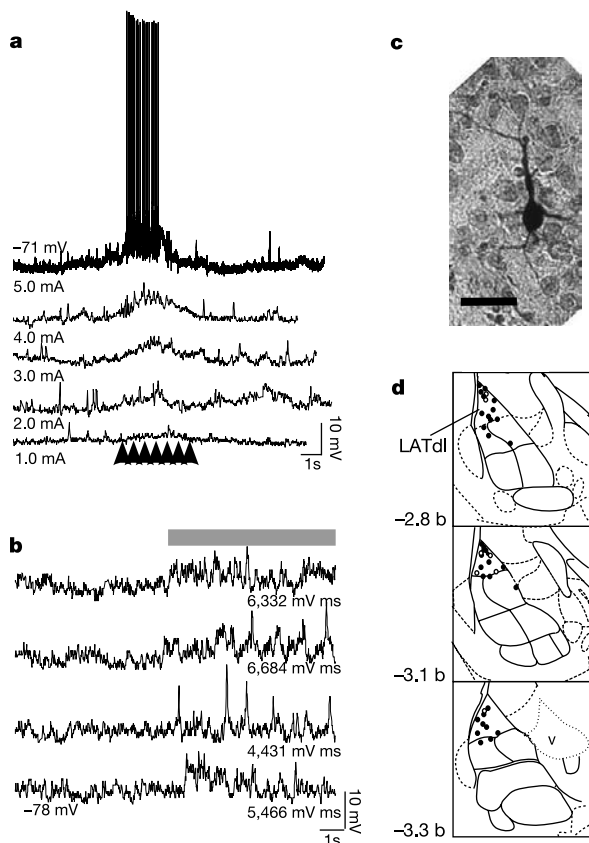


Figure 1 Stimulus-evoked responses in amygdala (LAT) neurons. **a**, Foot-shock (4 s, 20 Hz, arrows) evokes a stimulus intensity-dependent depolarizing response. **b**, Odour presentation (grey bar) results in a time-locked depolarization of the LAT neuronal membrane potential. The voltage traces represent responses of a single LAT neuron during two consecutive presentations of the odours anise (top traces) and almond (bottom traces). The area under the PSP is listed below each trace. **c**, Biocytin-filled LAT neurons display a morphology consistent with LAT projection neurons (scale bar, 40 μm). **d**, The majority of neurons were recorded from the dorsolateral region of the LAT (LATdl), as verified histologically. Open circles represent the location of recordings performed when haloperidol was administered. b, bregma; v, ventricle.

response to the first odour.

In addition to augmentation of the conditioned odour-evoked PSP, the pavlovian procedure resulted in enhanced neuronal excitability, as determined by intracellular current injection ($n = 16$, baseline mean $38.4 \pm 2.0 \text{ M}\Omega$; post-conditioning $52.5 \pm 3.6 \text{ M}\Omega$; $P < 0.01$, $t = -6.04$, d.f. = 15, paired t -test; Fig. 4). This change in neuronal excitability lasted for longer than 10 min when examined, and was also reflected in a reduction in the intracellular

positive current injection necessary for elicitation of an action potential (baseline $0.54 \pm 0.06 \text{ nA}$; post-conditioning $0.43 \pm 0.05 \text{ nA}$, $P < 0.05$, paired t -test), without concomitant changes in action potential threshold (baseline $-53.9 \pm 1.5 \text{ mV}$; post-conditioning $-56.1 \pm 1.9 \text{ mV}$, $P > 0.05$). Changes in neuronal excitability were also reflected in a shortening of the rise and decay of membrane time constants (τ_{rise} : baseline $11.6 \pm 0.4 \text{ ms}$; post-conditioning $8.6 \pm 0.6 \text{ ms}$; $P < 0.05$, paired t -test; τ_{decay} :

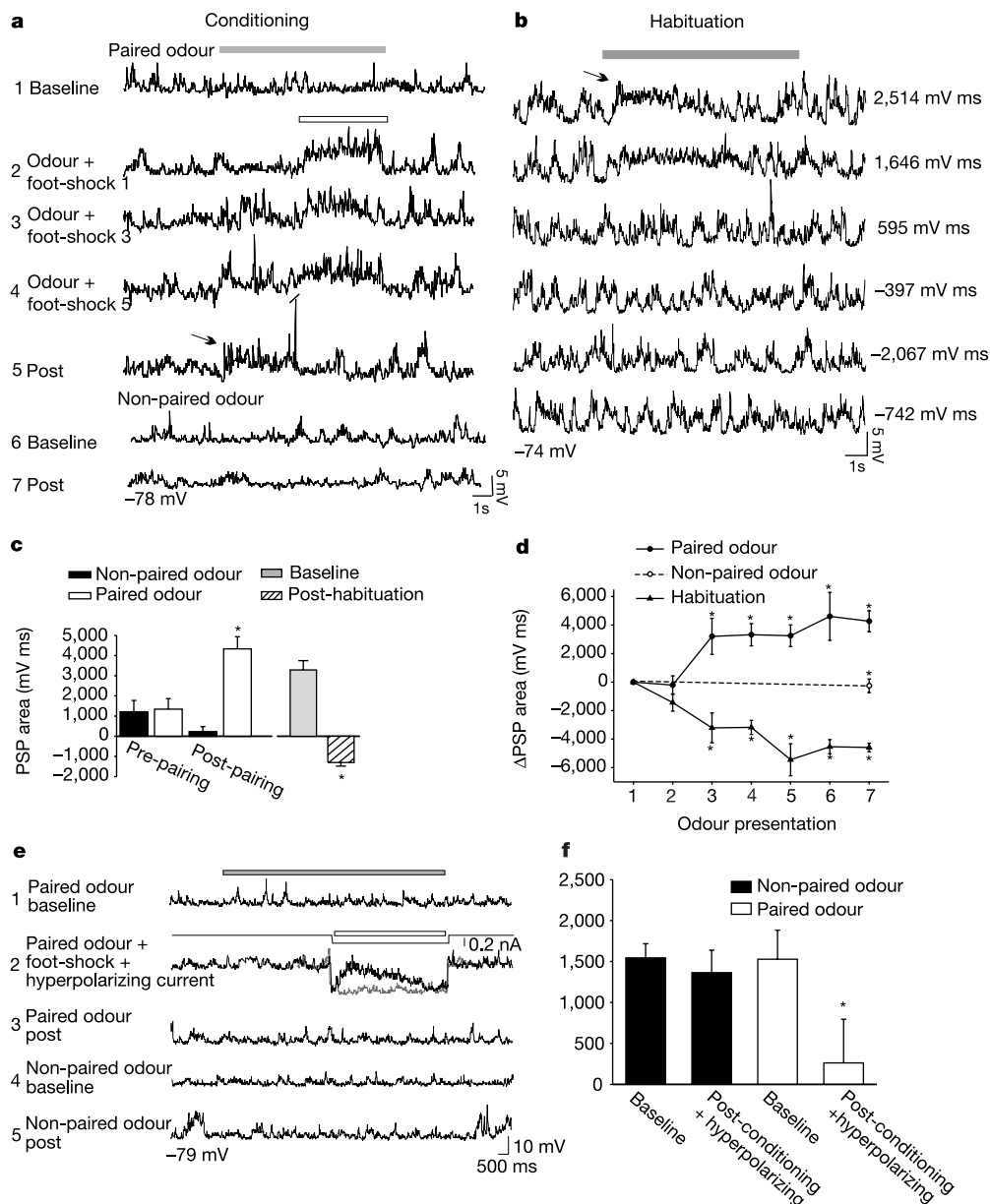


Figure 2 Opposing effects of habituation and conditioning on LAT neurons. **a**, Before pairing an odour with foot-shocks, the odour (grey bar) does not evoke a response in this LAT neuron (1). Throughout the course of pairings of the odour with the foot-shock (white bar) a response to this odour begins to emerge, as represented by enhanced deflections of the membrane potential (2–4). After conditioning, the paired odour alone evokes a response (5, arrow), whereas the response to the non-paired odour remains minimal (6, 7). All traces of odour-evoked responses show a 5-s epoch immediately before the odour, the odour presentation itself, and approximately 5 s after odour presentation. **b**, Repeated presentation of an odour (grey bar) in the absence of a foot-shock attenuates the odour-evoked response, quantified as area under PSP, on the right (arrow indicates baseline response to odour). **c**, **d**, Foot-shock-paired odours evoke significantly greater responses compared to their baseline response, and compared to the non-paired odour ($P < 0.01$ for both). In **d** the x axis represents the odour presentation procedure. Two consecutive

odour presentations are averaged for the baseline (1) and testing phases (7). The remainder are single odour presentations. After the baseline response was determined (after 1 on the x axis) the conditioning or habituation procedures were initiated. Odours presented without foot-shock ('habituation'; triangles) display a significant attenuation of response to below baseline-response amplitude, while paired odours (filled circles) display enhancement ($P < 0.01$ both). Non-paired control odours do not significantly change (dotted line, open circles). **e**, **f**, Addition of hyperpolarizing current (white bar, 2) during odour-foot-shock pairings negates the depolarization typically observed upon foot-shock administration and blocks enhancement of odour-evoked responses (3). The grey trace (2) represents the magnitude of hyperpolarization resulting from intracellular current injection in the absence of a foot-shock. A non-paired odour also does not display enhancement (4, 5).

Table 1 Effect of haloperidol on odour-evoked responses

	Baseline odour response (mV ms)	Post conditioning 1 (mV ms)	Post conditioning 2 (mV ms)
Saline ($n=4$)			
First odour paired	57.9 ± 539.3	$2,174.1 \pm 323.9^{*†}$	$2,748.5 \pm 363.8^{*}$
Second odour paired	234.7 ± 755.7	-105.6 ± 232.5	$4,229.8 \pm 320.8^{*}$
Haloperidol ($n=7$)			
First odour paired	154.6 ± 621.4	$4,036.4 \pm 491.6^{*†}$	$3,117.2 \pm 252.7^{*†}$
Second odour paired	104.4 ± 928.6	-473.2 ± 345.8	-441.9 ± 367.2

*Significant difference from baseline odour response, paired t -test, $P < 0.05$.

†Significant difference from response to other odour, t -test $P < 0.05$.

baseline 13.2 ± 0.7 ms; post-conditioning 10.4 ± 0.8 ms; $P < 0.05$, paired t -test). When saline was administered and the second odour was paired with a foot-shock, an additional increase in neuronal excitability was observed ($n = 4$, mean 62.4 ± 10.9 M Ω ; $P = 0.016$, $t = -37.6$, d.f. = 2, paired t -test; Fig. 4).

This increase in neuronal excitability was not observed when haloperidol was administered instead of saline ($n = 6$, mean 41.9 ± 5.0 M Ω , $P > 0.05$, paired t -test, Fig. 4). Haloperidol not only prevents the increased input resistance due to the second conditioning procedure, resulting in a significant difference between saline- and haloperidol-treated groups ($P < 0.05$, $t = 2.36$, d.f. = 8, t -test), but it also reduces the input resistance, restoring it back to pre-conditioning baseline levels ($P = 0.09$,

$t = -2.08$, d.f. = 5, paired t -test). This leads to the possibility that enhanced post-synaptic neuronal excitability mediated by dopamine receptor activation may be necessary for pavlovian conditioning. These data demonstrate that blockade of dopamine receptors during conditioning attenuates the neuronal plasticity of LAT neurons, indicating that dopamine plays a role in the initial phases of affective conditioning, and is not involved solely in consolidation phases. Furthermore, dopamine may impose its actions via enhancement of LAT neuronal excitability, thereby augmenting conditioned odour-evoked responses. However, heightened tonic dopamine levels alone do not appear to be sufficient to induce this plasticity, as even relatively high doses of the dopamine agonist apomorphine (1.3 ± 0.03 mg kg $^{-1}$

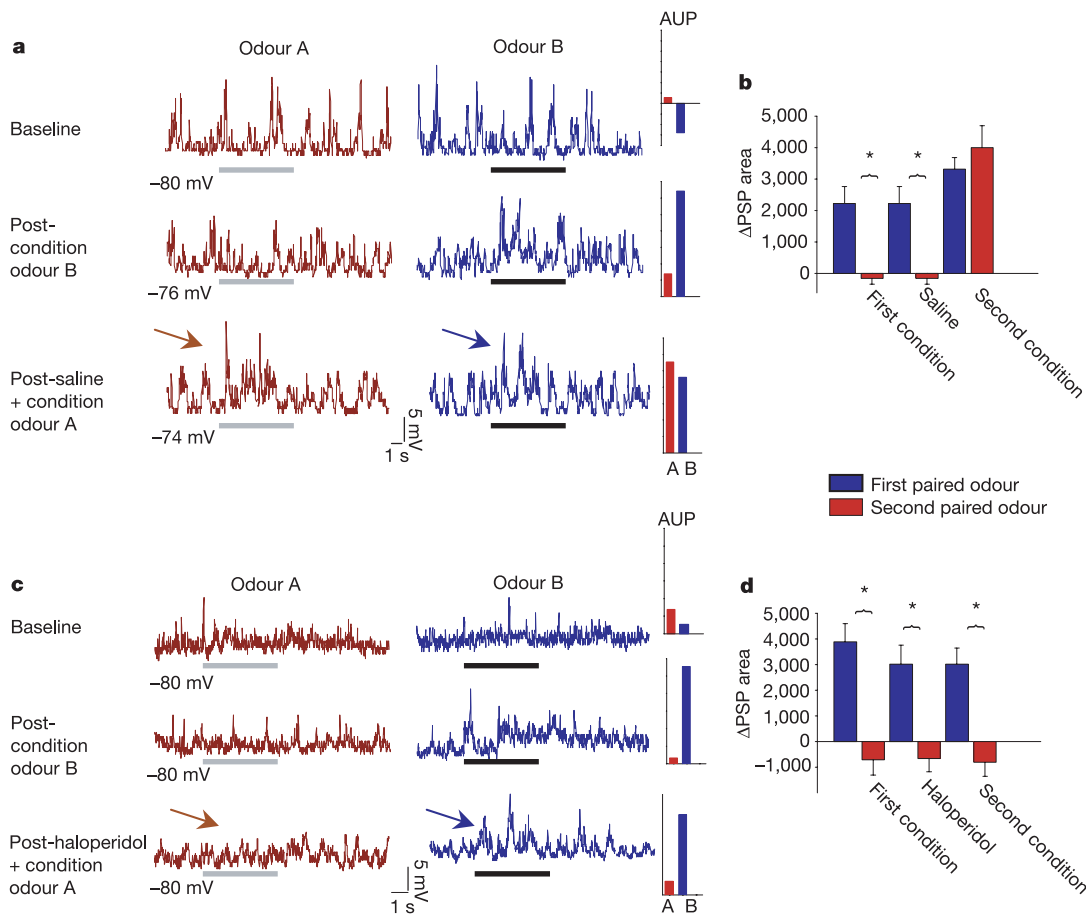


Figure 3 Dopamine receptor antagonism blocks enhancement of the odour-evoked response caused by the conditioning procedure. **a, b**, Neither odour presentation evokes a large time-locked PSP relative to baseline spontaneous PSPs (traces show first 5 s of odour presentation represented by solid bars in every trace). This is also demonstrated by the bar graphs to the right of each trace, in which the amplitude of the evoked response to odour A (red bar) and odour B (blue bar) is illustrated (x-axis length = 7,000 mV ms). After conditioning of odour B, this odour now evokes a large PSP (middle traces). When saline is administered and odour A is paired with a foot-shock, both odour A and odour B now

evoke large PSPs (bottom trace, red and blue arrows, respectively). This can be quantified by area under the PSP. **c, d**, In another neuron, odour A and odour B both evoke small PSP responses (the top trace shows odours presented at solid bars). After pairing of odour B with a foot-shock, this odour now evokes a large PSP (middle trace). After haloperidol administration (0.67 mg kg $^{-1}$ i.v.) odour B still evokes a PSP (blue arrow); however, odour A fails to evoke a PSP after pairing of odour A with foot-shock (red arrow). Upon quantification by area under PSP, significant differences between the two odours are observed throughout the experiment.

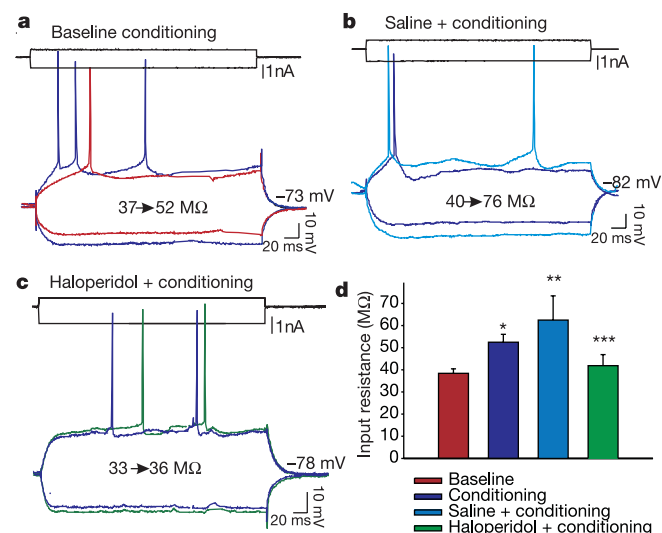


Figure 4 Neuronal excitability is enhanced by the conditioning procedure, which is blocked by dopamine receptor antagonism. **a**, During conditioning, the input resistance of this neuron increases from 37 MΩ to 52 MΩ, and more action potentials are evoked by the same amplitude current pulse (dark blue trace). **b**, When saline is administered and the conditioning procedure is repeated with the second odour, the response to current injection is further enhanced (dark blue trace, following first conditioning; light blue trace, following saline + second conditioning procedure). **c**, Haloperidol administration blocks the enhanced neuronal excitability usually resulting from the conditioning of the second odour (dark blue trace, following first conditioning procedure; green trace, following haloperidol + second conditioning procedure). **d**, Conditioning significantly enhances neuronal excitability, which is further enhanced by a second conditioning procedure. This enhancement is blocked, and partially reversed upon administration of haloperidol. ($P < 0.05$; asterisk, significantly different from baseline input resistance; double asterisk, significantly different from baseline and post-conditioning input resistance; triple asterisk, significantly different from saline + conditioning.)

intravenously, $n = 4$) did not significantly enhance odour-evoked PSPs (baseline $2,045.5 \pm 743.1$ mV ms; post-apomorphine $2,244.9 \pm 684.9$ mV ms; $P > 0.05$, $t = -0.76$, d.f. = 7, paired t -test).

It appears that the temporal association of the conditioned odour with the foot-shock itself is necessary to augment the response of this odour (as observed in behavioural studies), as a general enhancement of neuronal excitability itself would also be expected to enhance non-paired odours. It may be that a combination between pairing of the odour with post-synaptic depolarization resulting from an unconditioned stimulus (that is, the foot-shock) in the presence of dopamine is required for the augmentation to the paired odour. The activation of a set of inputs representative of one odour in concert with a post-synaptic depolarization resulting from foot-shock may confer selectivity upon the response to this odour, and not to other odours that are represented by another combination of inputs. Thus, enhanced neuronal excitability in itself may not be the primary variable responsible for the specific enhancement of paired odours, which may take place at the synapse. Instead, enhanced neuronal excitability may help in conferring this enhancement, perhaps by enhancement of foot-shock-evoked depolarizations, or by supralinear enhancement of odour-evoked potentials that occur at the same time as the foot-shock.

Previous studies have demonstrated that dopamine receptor activation enhances the excitability of LAT neurons, as well as reducing baseline spontaneous membrane fluctuation, as represented by the standard deviation (s.d.) of the membrane potential²². Therefore, the s.d. was measured to further assess the role of dopamine in conditioning. The conditioning procedure led to a 27% decrease in the s.d. of the membrane potential (baseline mean 2.7 ± 0.3 mV; post-conditioning 2.0 ± 0.2 mV; $P < 0.01$, $t = 6.35$,

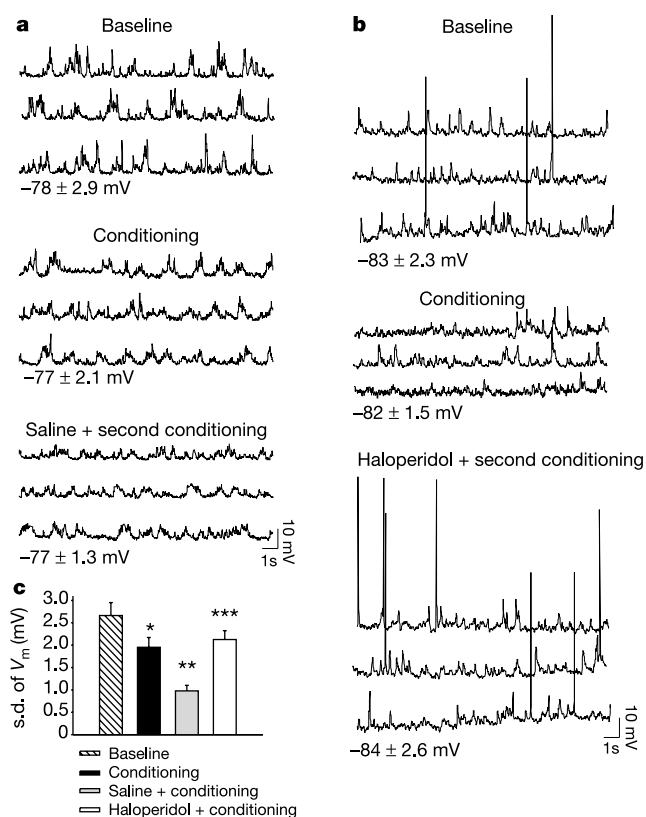


Figure 5 The pavlovian conditioning procedure reduces spontaneous PSPs via dopamine receptor activation. **a**, The pavlovian conditioning procedure reduces spontaneously occurring PSPs (middle set of traces compared to top set of traces). This is further reduced by a second conditioning procedure (bottom set of traces). **b**, Haloperidol administration in place of saline blocks the expected decrement of spontaneously occurring PSPs usually observed after the second conditioning. **c**, The alterations of spontaneous PSPs during conditioning can be quantified as the standard deviation of the mean membrane potential. Haloperidol administration blocks the effects of conditioning on spontaneous PSPs. ($P < 0.05$; asterisk, significant difference from baseline s.d.; double asterisk, significant difference from baseline and saline + conditioning; triple asterisk, significant difference from saline + conditioning.) V_m, membrane potential.

d.f. = 24, paired t -test; Fig. 5). Similarly to the above, administration of saline and repetition of the conditioning procedure on the second odour led to a further decrease in the s.d. (1.0 ± 0.1 mV; $P = 0.025$, $t = 3.52$, d.f. = 4, paired t -test; Fig. 5). Administration of haloperidol instead of saline blocked the reduction in s.d. induced by the second conditioning procedure and partly reversed the effects of the initial conditioning (2.1 ± 0.2 mV; $P < 0.01$, $t = -6.38$, d.f. = 6, paired t -test; Fig. 5). This also resulted in a significant difference of the s.d. of the membrane potential between saline- and haloperidol-treated rats ($P < 0.01$, $t = -4.40$, d.f. = 10, t -test). We notice that enhanced input resistance is not associated with increased fluctuations of the membrane potential, as measured by s.d. It is possible that the reduced s.d. is reflective of decreased dendritic signal propagation or a pre-synaptic effect not measured in these studies, resulting in less baseline fluctuation, which thereby enhances signal fidelity²².

We have demonstrated that a pavlovian conditioning procedure using pairing of natural stimuli can induce neuronal plasticity in the LAT of anaesthetized rats, a region known to be critical in the formation of affective associations^{1–11}. Moreover, the plasticity and alterations of neuronal excitability due to the conditioning procedure are blocked by a dopamine antagonist. Indeed, our data show that dopamine is the critical link between synaptic drive and membrane excitability that underlies associative events. This is consistent with the role of dopamine in amygdala-dependent

affective responses in behaving animals^{12–15}. Thus, one consequence of pharmacological intervention in dopamine system dynamics is a disruption in this finely tuned associative process. Such a condition may underlie improper associative events, such as those thought to contribute to inappropriate affect in schizophrenia^{23–25} or heightened distractibility in attention-deficit hyperactivity disorders^{26,27}. □

Methods

Materials

Haloperidol was a gift from McNeil Laboratories, and was dissolved in dilute lactic acid, then further diluted with 0.9% saline to a concentration of 0.5 mg ml⁻¹.

Animal preparation

All procedures were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals, and were approved by the University of Pittsburgh Institutional Animal Care and Use Committee. Male Sprague–Dawley rats (250–350 g weight) were anaesthetized with an initial injection of 400 mg kg⁻¹ of 8% chloral hydrate, administered intraperitoneally, and placed into a stereotaxic device that was modified to allow contained odour delivery to the nose at a flow rate of 1.5 l min⁻¹, with delivery of odour within 0.5 s. Supplemental anaesthesia (8% chloral hydrate) was delivered by a lateral tail vein catheter as necessary to maintain suppression of hindlimb withdrawal reflex. Temperature was monitored and maintained at ~37°C. Coordinates for recordings were determined using a stereotaxic atlas, as follows: LAT –5.0 lateral, –3.3 caudal from bregma.

Intracellular recordings

Recordings were performed as described previously²². Briefly, electrodes were constructed using borosilicate glass tubing and filled with 2% biocytin in 3 M potassium acetate (Sigma). Impedances were measured *in situ* and ranged from 45–75 MΩ. Hyperpolarizing d.c. pulses were used to determine input resistance, and only the linear portion of the plot was included for this analysis. Mean resting membrane potential and standard deviation were determined from 30-s sampling periods, and action potentials were eliminated from this analysis. The area under the odour-evoked PSP was analysed by measuring the first 5 s of the area under the odour-evoked PSP as the baseline, and subtracting out the area under spontaneously occurring PSPs in the 5 s immediately preceding odour presentation. These baseline-corrected responses to odours were averaged for data analysis. Recording electrode placements were identified as previously described²². Neurons were not included in this study if the resting membrane potential was less polarized than –65 mV, if their action potentials did not overshoot 0 mV, if the measured input resistance was below 20 MΩ, or if they were found to lie outside the LAT.

Pavlovian conditioning

A pavlovian conditioning procedure was performed by pairing of an odour with a foot-shock. A foot-shock was delivered by two 28-g needles inserted in the lateral side of the foot contralateral to the neuronal recordings. Each odour (anise or almond) was presented at least two times, for 10 s, with a 60-s delay between presentations. One odour was chosen to be paired with the foot-shock. Paired odour selection was counterbalanced. This first odour was paired with the foot-shock (4 s, 2–5 mA, 20 Hz, 0.2-ms duration pulses) such that the foot-shock was presented 5 s after the odour began. The foot-shock intensity chosen was dependent on the level of depolarization achieved in the neuron by the foot-shock. Typically, an intensity of 4–5 mA was chosen, evoking a response that was subthreshold to spike generation. This pairing was performed 5–8 times at 60-s intervals. After these pairings, each odour was presented at least twice, 10 s each, at 60 s intervals. In some experiments, after the first odour was paired with the foot-shock, 0.9% saline (0.4 ml) or haloperidol (0.4 ml of 0.5 mg ml⁻¹) was administered, and the rat was subjected to exactly the same procedure, substituting the non-paired odour for the previously paired odour.

To examine habituation, the odours were presented to a separate group of rats in a fashion identical to the conditioning protocol, but without foot-shocks. Only neurons that displayed an initial response to odour presentation were used to examine habituation.

Received 29 January; accepted 4 March 2002.

1. Maren, S. Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends Neurosci.* **22**, 561–567 (1999).
2. Davis, M., Rainnie, D. & Cassell, M. Neurotransmission in rat amygdala related to fear and anxiety. *Trends Neurosci.* **17**, 208–214 (1994).
3. Cousens, G. & Otto, T. Both pre- and posttraining excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. *Behav. Neurosci.* **112**, 1092–1103 (1998).
4. Faneslow, M. S. & LeDoux, J. E. Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* **23**, 229–232 (1999).
5. Ben-Ari, Y. & Le Gal La Salle, G. Plasticity at the unitary level. II. Modifications during sensory-related association procedures. *Electroencephalogr. Clin. Neurophysiol.* **32**, 667–679 (1972).
6. Maren, S., Poremba, A. & Gabriel, M. Basolateral amygdaloid multi-unit neuronal correlates of discriminative avoidance learning in rabbits. *Brain Res.* **549**, 311–316 (1991).
7. Rogan, M. T., Staubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
8. Nishijo, H., Uwano, T., Tamura, R. & Ono, T. Gustatory and multimodal neuronal responses in the amygdala during licking and discrimination of sensory stimuli in awake rats. *J. Neurophys.* **79**, 21–36 (1998).

9. Pare, D. & Collins, D. R. Neuronal correlates of fear in the lateral amygdala: multiple extracellular recordings in conscious cats. *J. Neurosci.* **20**, 2701–2710 (2000).
10. Collins, D. R. & Pare, D. Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the CS+ and CS–. *Learn. Mem.* **7**, 97–103 (2000).
11. Repa, J. C. *et al.* Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neurosci.* **4**, 724–731 (2001).
12. Guarraci, F. A., Frohardt, R. J. & Kapp, B. S. Amygdaloid D1 dopamine receptor involvement in Pavlovian fear conditioning. *Brain Res.* **827**, 28–40 (1999).
13. Guarraci, F. A., Frohardt, R. J., Falls, W. A. & Kapp, B. S. The effects of intraamygdaloid infusions of a D2 dopamine receptor antagonist on Pavlovian fear conditioning. *Behav. Neurosci.* **114**, 647–651 (2000).
14. Greba, Q. & Kokkinidis, L. Peripheral and intraamygdalar administration of the dopamine D1 receptor antagonist SCH23390 blocks fear-potentiated startle but not shock reactivity or the shock sensitization of acoustic startle. *Behav. Neurosci.* **114**, 2, 262–272 (2001).
15. Greba, Q., Gifkins, A. & Kokkinidis, L. Inhibition of amygdaloid dopamine D2 receptors impairs emotional learning measured with fear-potentiated startle. *Brain Res.* **899**, 218–226 (2001).
16. Rescorla, R. A. Pavlovian conditioning and its proper control procedures. *Psychol. Rev.* **74**, 1, 71–80 (1967).
17. Dutton, R. C. *et al.* The concentration of isoflurane required to suppress learning depends on the type of learning. *Anesthesiology* **94**, 514–519 (2001).
18. Gold, P. E., Weinberger, N. M. & Sternberg, D. B. Epinephrine-induced learning under anesthesia: Retention performance at several training-testing intervals. *Behav. Neurosci.* **99**, 1019–1022 (1985).
19. Ghoneim, M. M. & Block, R. I. Learning and memory during general anesthesia: An update. *Anesthesiology* **87**, 387–410 (1997).
20. Weinberger, N. M., Gold, P. E. & Sternberg, D. B. Epinephrine enables Pavlovian fear conditioning under anesthesia. *Science* **223**, 605–607 (1984).
21. Sacchetti, B., Lorenzini, C. A., Baldi, E., Tassoni, G. & Bucherelli, C. Auditory thalamus, dorsal hippocampus, basolateral amygdala, and perirhinal cortex role in the consolidation of conditioned freezing to context and to acoustic conditioned stimulus in the rat. *J. Neurosci.* **19**, 21, 9570–9578 (1999).
22. Rosenkranz, J. A. & Grace, A. A. Cellular mechanisms of infralimbic and prelimbic prefrontal cortical inhibition and dopaminergic modulation of basolateral amygdala neurons in vivo. *J. Neurosci.* **22** (1), 324–337 (2002).
23. Andreasen, N. C. *et al.* Positive and negative symptoms of schizophrenia: past, present, and future. *Acta Psychiatrica Scan.* (Suppl.) **384**, 51–59 (1994).
24. Morrison, R. L., Bellack, A. S. & Mueser, K. T. Deficits in facial-affect recognition and schizophrenia. *Schiz. Bull.* **14** (1), 67–83 (1988).
25. Grossberg, S. The imbalanced brain: from normal behaviour to schizophrenia. *Biol. Psychiatry* **48**, 81–98 (2000).
26. Nigg, J. T. Is ADHD a disinhibitory disorder? *Psychol. Bull.* **127**, 571–598 (2001).
27. McGough, J. J. & McCracken, J. T. Assessment of attention deficit hyperactivity disorder: a review of recent literature. *Curr. Opin. Ped.* **12**, 319–324 (2000).

Acknowledgements

We thank N. MacMurdo, C. Smolak and B. Lowry for technical assistance, and H. Moore, A. R. West and S. Stocker for discussions. Support was provided by National Institutes of Health grants.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.R. (e-mail: rosenk@bns.pitt.edu).

Stress response genes protect against lethal effects of sleep deprivation in *Drosophila*

Paul J. Shaw*, Giulio Tononi†*, Ralph J. Greenspan* & Donald F. Robinson*

* The Neurosciences Institute, 10640 John J. Hopkins Drive, San Diego, California 92121, USA

† Present addresses: University of Wisconsin-Madison, Department of Psychiatry, 6001 Research Park Boulevard, Madison, Wisconsin 53711, USA

Sleep is controlled by two processes: a homeostatic drive that increases during waking and dissipates during sleep, and a circadian pacemaker that controls its timing¹. Although these two systems can operate independently^{2,3}, recent studies indicate a more intimate relationship^{4,5}. To study the interaction between homeostatic and circadian processes in *Drosophila*, we examined

homeostasis in the canonical loss-of-function clock mutants *period* (*per*⁰¹), *timeless* (*tim*⁰¹), *clock* (*Clk*^{*irk*}) and *cycle* (*cyc*⁰¹)^{6–9}. *cyc*⁰¹ mutants showed a disproportionately large sleep rebound and died after 10 hours of sleep deprivation, although they were more resistant than other clock mutants to various stressors. Unlike other clock mutants, *cyc*⁰¹ flies showed a reduced expression of heat-shock genes after sleep loss. However, activating heat-shock genes before sleep deprivation rescued *cyc*⁰¹ flies from its lethal effects. Consistent with the protective effect of heat-shock genes, was the observation that flies carrying a mutation for the heat-shock protein *Hsp83* (*Hsp83*⁰⁸⁴⁴⁵)¹⁰ showed exaggerated homeostatic response and died after sleep deprivation. These data represent the first step in identifying the molecular mechanisms that constitute the sleep homeostat.

A sleep-like state has been described in *Drosophila melanogaster* on the basis of its similarities to mammalian sleep^{11,12}. This state is characterized by increased arousal thresholds and is regulated homeostatically^{11,12}. Like mammalian sleep, it is abundant in young flies, decreases in older animals and is modulated by stimulants and hypnotics¹¹. Perhaps the most important similarity between mammals and flies is homeostatic regulation: when flies are kept awake, they show a large compensatory increase in sleep the

next day^{11,12}.

In mammals, the circadian pacemaker alternately promotes and maintains both wakefulness and sleep^{13,14}. Although the circadian pacemaker and the sleep homeostat can interact, little is known about the underlying mechanisms. To evaluate this relationship, homeostasis was evaluated in clock mutants maintained in constant darkness (DD) and deprived of sleep for 3, 6, 9 and 12 h (Fig. 1). Under these conditions, sleep is evenly distributed across the day (Fig. 1a). Upon release from sleep deprivation, wild-type *Canton-S* (*Cs*) flies recover ~30–40% of the sleep that they lost within 12 h (ref. 11). *per*⁰¹ and *Clk*^{*irk*} showed a more prominent sleep rebound, reclaiming ~100% of lost sleep within 12 h (Fig. 1a–e). *tim*⁰¹ flies did not show a homeostatic response after 3–6 h of sleep deprivation¹² but displayed a sleep rebound similar to that of *per*⁰¹ and *Clk*^{*irk*} flies after 7, 9 and 12 h of sleep deprivation (*P* > 0.05; 7-h data not shown).

Surprisingly, *cyc*⁰¹ mutants showed an exaggerated response to 3 h of sleep deprivation, reclaiming ~3 min of sleep over baseline for each minute of sleep lost (Fig. 1b). Further increasing sleep debt produced a change in the regulation of sleep not seen in other clock mutants (Fig. 1c–e): *cyc*⁰¹ flies showed large increases in sleep that persisted for as long as the flies were recorded (up to 16 days). These

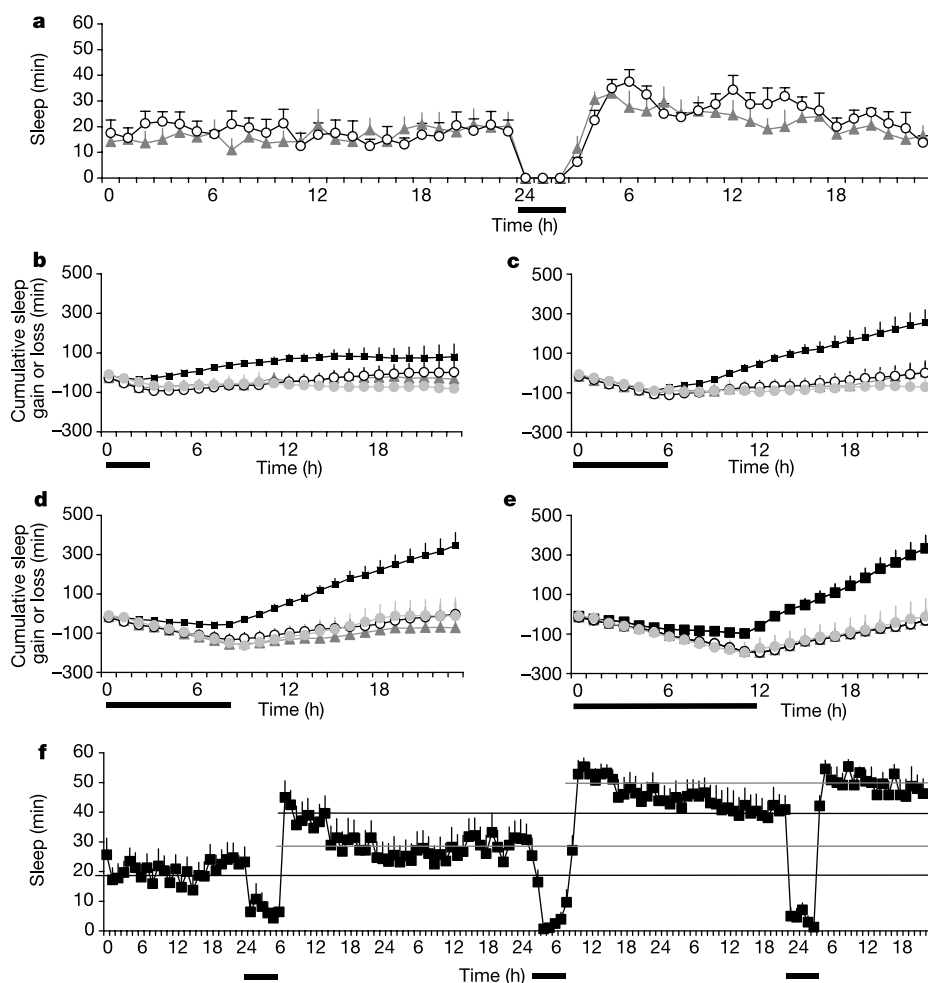


Figure 1 Sleep homeostasis is altered in circadian mutants and is markedly increased in *cyc*⁰¹ flies. **a**, Illustration of sleep during a typical experiment. Undisturbed *per*⁰¹ and *Clk*^{*irk*} flies sleep for ~20 min each hour under DD. When flies are kept awake for 3 h they show an initial increase in sleep followed by a return to baseline. **b**, Cumulative sleep lost or gained during sleep deprivation and subsequent recovery. A negative slope indicates sleep lost, a positive slope indicates sleep gained; when the slope is zero, recovery is

complete. **c–e**, *cyc*⁰¹ flies show a disproportionate increase in sleep that depends on the length of the deprivation. **f**, *cyc*⁰¹ flies continually increase baseline levels of sleep after repeated sleep deprivation. Black squares, *cyc*⁰¹; grey triangles, *Clk*^{*irk*}; open circles, *per*⁰¹; grey circles, *tim*⁰¹; horizontal lines reflect new setpoint; black bars indicate periods of sleep deprivation.

Table 1 Mortality in clock mutants during 12 h of sleep deprivation

Genotype	No. of trials	<i>n</i>	Mortality (%)	Range (%)
<i>cyc⁰¹</i>	16	251	33 ± 4	12–60
<i>Clk^{irk}</i>	9	138	0	n.a.
<i>per⁰¹</i>	7	106	0	n.a.
<i>tim⁰¹</i>	5	87	0	n.a.
Cs	20	540	0	n.a.

periods of quiescence were associated with increased arousal thresholds, indicating that the deprivation had produced an increase in the amount of sleep and did not result in an injured fly. If sleep deprivation produced an increased need for sleep in *cyc⁰¹* flies, they should show higher amounts of sleep when sleep deprived for a second time. Indeed, deprivations of an additional 6 h resulted in further increases in sleep (Fig. 1f).

The extreme sensitivity of *cyc⁰¹* flies was revealed when sleep deprivation extended past 10 h: the flies began to die. This effect was not observed in wild-type flies, in mutant lines representing 45 other genetic loci, or in other clock mutants, indicating that the mutations do not in themselves increase vulnerability to sleep deprivation (Table 1, and Supplementary Information). Note that the most sensitive of the clock mutants (*cyc⁰¹*) is the only one that does not cycle.

To determine whether death in the *cyc⁰¹* flies was due to the stimuli used for sleep deprivation, flies were deprived of sleep for 30 min each hour for 24 h, ensuring that the flies received the same number of stimuli that accrued during 12 h of sleep deprivation but without producing 12 h of continuous wakefulness. No deaths were observed after this protocol, indicating that the deprivation stimulus was not responsible for the deaths (Fig. 2a). Further supporting this conclusion, *stress-sensitive B* (*sesB¹*) flies, which are extremely sensitive to mechanical shock¹⁵, survived 12 h of sleep deprivation and showed activity patterns during the deprivation that were similar to those of wild-type flies ($P = 0.36$, data not shown). We also deprived *cyc⁰¹* flies of sleep by gentle handling as described previously¹¹. The proportion of flies that succumbed to sleep deprivation and the size of the homeostatic response in surviving flies were indistinguishable from the automated deprivation method (two trials, $n = 32$, $P = 0.67$). Similar results were obtained

with a rotating deprivation apparatus described previously⁸.

To determine whether death in *cyc⁰¹* flies is due specifically to sleep deprivation or to hypersensitivity to any environmental challenge, *per⁰¹*, *tim⁰¹*, *Clk^{irk}*, *cyc⁰¹* and Cs flies were subjected to several stressors including heat stress, oxidative stress, starvation, desiccation and physical stress. *cyc⁰¹* flies were as sensitive, but no more so than other genotypes to desiccation and vortex-mixing (Fig. 2e, f; $P > 0.05$) and survived longer than *per⁰¹*, *tim⁰¹* and *Clk^{irk}* flies when challenged with heat, oxidative stress and starvation (Fig. 2b–d; $P < 0.01$). Cs flies, which have an intact clock, were more resistant to starvation and desiccation than *tim⁰¹*, *Clk^{irk}* and *cyc⁰¹* flies. These data indicate that *cyc⁰¹* mutants are vulnerable to prolonged wakefulness in itself and are not merely hypersensitive to non-related stressors.

To confirm that this phenotype maps to the *cyc* locus, we crossed *cyc⁰¹* homozygotes with flies carrying the appropriate deficiency *Df(3L)kto2/TM6B, Tb¹*. The resulting *cyc⁰¹/Df* transheterozygote flies showed an exaggerated homeostatic response and deaths after 12 h of sleep deprivation (data not shown). Furthermore, *cyc⁰¹* heterozygotes with and without a functioning clock (*cyc⁰¹/+* and *per⁰¹;cyc⁰¹/+*) also showed exaggerated homeostasis (data not shown). We evaluated *cyc⁰¹ st* and *yw;cyc⁰¹* flies to determine whether the background would influence the phenotype; it did not do so (data not shown). Nor was the phenotype changed in aged flies (25 days old; data not shown). Similarly to females, male *cyc⁰¹* flies have been shown to have aberrations in sleep homeostasis (J. Hendricks, personal communication). We found that *cyc⁰¹* males were less sensitive to stress than other clock mutant males, recovered 100% of lost sleep (compared with 300% in *cyc⁰¹* females) and died after 12 h of sleep deprivation, indicating that homeostasis is dissociable from lethality (data not shown). Interestingly, homeostatic regulation of sleep is also sex-dependent in humans¹⁶.

Prolonged sleep deprivation (2–4 weeks) is invariably fatal in normal rats¹⁷. Is the rapid demise after a few hours of sleep deprivation the result of an anomalous reaction in *cyc⁰¹* mutants, or is it an increased susceptibility to the lethal consequences of sleep loss? Individual Cs flies were kept awake for 70 h by tapping on their tubes when they stopped moving, as described above, to ensure that lethality was not due to excessive handling; 2 of 12 Cs flies died after ~60 h of continuous wakefulness, whereas 2 more died by ~67–

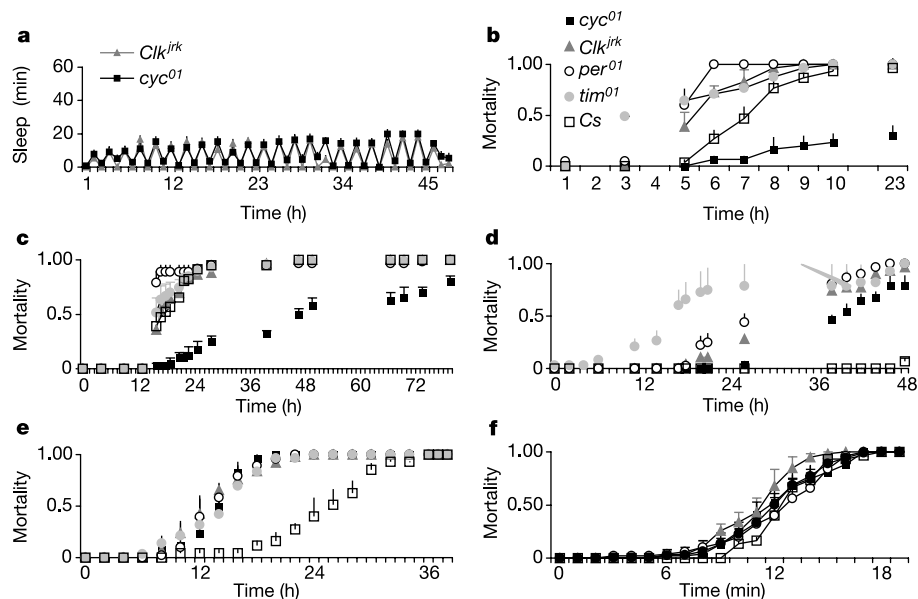


Figure 2 *cyc⁰¹* flies are resistant to stress. **a**, The amount of sleep in *cyc⁰¹* and *Clk^{irk}* flies deprived of sleep for 30 min each hour. **b–f**, Mortality in *cyc⁰¹*, *Clk^{irk}*, *per⁰¹* and

tim⁰¹ flies in response to heat stress (36 °C) (**b**), oxidative stress (20 μM paraquat) (**c**), starvation (**d**), desiccation (**e**) and physical stress (vortex-mixing) (**f**).

70 h. The behaviour of the flies during the last hours of the sleep deprivation protocol resembled that seen in *cyc⁰¹* flies, indicating that the deaths were due to sleep loss and not to the deprivation stimulus itself. To test this, we kept an additional group of *Cs* flies ($n = 10$) awake by using a different deprivation method and found again that flies began to die between 60 and 70 h. These data indicate that sleep does indeed serve a vital biological role in the fly and that specific mutations that increase susceptibility to death might help to clarify such a role.

Given that *cyc⁰¹* flies are equally well or better equipped than other clock mutants to tolerate chronic heat and other stressors, why do they die in response to sleep deprivation? There is much evidence that stress response genes can protect an organism during challenging conditions¹⁸. We therefore examined the ability of heat and sleep deprivation to activate stress response genes, by using real-time quantitative polymerase chain reaction (QPCR). All clock mutants responded to 3 h of heat with an induction of genes such as *hsp70*, *Hsp83*, *droj1* and *hsc70-3* coding for chaperone proteins (Fig. 3a)^{19–21}. After 3 h of sleep deprivation, the levels of these genes were

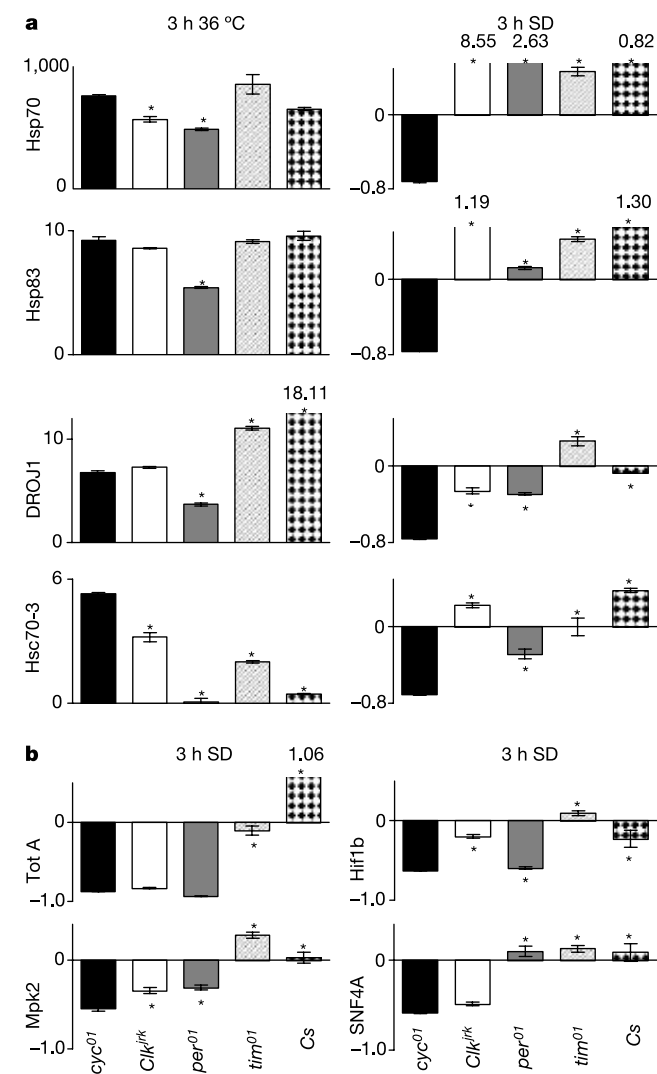


Figure 3 Expression of heat-shock genes is reduced in *cyc⁰¹* flies after 3 h of sleep deprivation. **a**, Change in gene expression after either 3 h of heat exposure (left panels) or 3 h of sleep deprivation (SD) (right panels) shown as percentage deviation from control values. **b**, Expression of stress response genes after 3 h (SD). Values for data extending beyond the axis are shown above respective bars. Asterisks indicate values that are significantly different from *cyc⁰¹* ($P < 0.05$).

near baseline in *Clk^{rk}* flies (with the exception of *hsp70*) and were unchanged or mildly increased in *per⁰¹*, *tim⁰¹* and *Cs* flies (see Supplementary Information for gene expression profiles after longer deprivation protocols). Interestingly, levels of chaperone proteins are also elevated after sleep deprivation in rodents²². However, sleep deprivation produced a decrease in the expression of these genes in *cyc⁰¹* flies. Genes activated by qualitatively different stressors, including metabolic stress (*SNF4a*, *Hif1*), chemical stress (*mpk2*) and humoral stress (*turandot*), were reduced in all lines^{23–26}, indicating that sleep deprivation is not inherently stressful (Fig. 3b).

To evaluate the relationship between heat-shock genes and sleep deprivation in *cyc⁰¹* flies, we induced heat-shock genes before sleep deprivation. When 12 h of sleep deprivation was preceded by 3 h of heat exposure at 36 °C, the mortality rate was reduced compared to unheated *cyc⁰¹* flies (Fig. 4a; note that one would have predicted increased mortality because preheating results in a further 3 h of wakefulness). Moreover, heat exposure reduced homeostatic drive in *cyc⁰¹* and *Clk^{rk}* flies (Fig. 4b; *Clk^{rk}* data not shown). When *cyc⁰¹* flies were pre-exposed to 37 °C, homeostasis was reduced further (Fig. 4b).

We also show the importance of heat-shock genes in sleep deprivation by examining flies mutant for *Hsp83* (*Hsp83⁰⁸⁴⁴⁵*)¹⁰. After 12 h of sleep deprivation, *Hsp83⁰⁸⁴⁴⁵* mutants exhibited a mortality rate similar to that of *cyc⁰¹* and showed a homeostatic response corresponding to fivefold that of wild-type flies (three trials; $n = 48$; Fig. 4c, d). The sensitivity to sleep deprivation in *Hsp83⁰⁸⁴⁴⁵* mutants is present even in heterozygous flies, which have only a modest reduction in gene product. Heterozygous *Hsp83⁰⁸⁴⁴⁵* flies displayed a sleep rebound that was not statistically different from either homozygous *Hsp83⁰⁸⁴⁴⁵* or heterozygous *Hsp83^{c6A}* flies ($P > 0.10$; data not shown). However, both *Hsp83* heterozygotes exhibited a sleep rebound that was significantly different from that of *Hsp60^{RA75}* heterozygotes²⁷, indicating that a limited set of chaperone proteins are involved in homeostasis ($P < 0.05$; Fig. 4d). Finally, whereas preheating *cyc⁰¹* flies prevented the lethal effects of sleep deprivation, this did not occur in *Hsp83⁰⁸⁴⁴⁵* flies (Fig. 4c). It should be noted that we have evaluated sleep homeostasis in mutant lines representing 45 other genetic loci; *cyc⁰¹* and

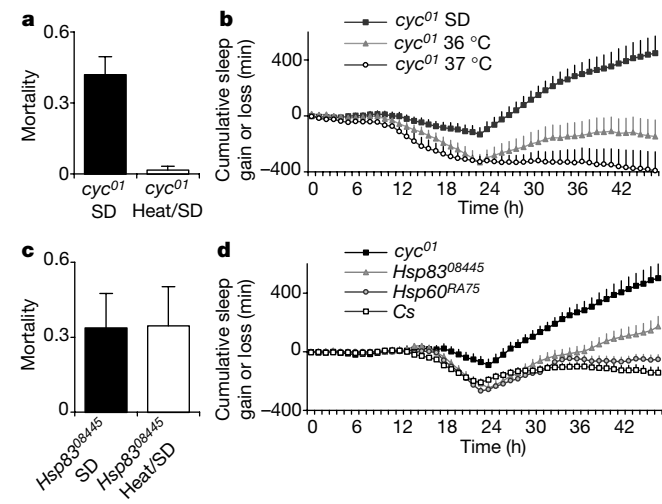


Figure 4 Heat protects *cyc⁰¹* flies from the lethal effects of sleep deprivation. **a**, Mortality after 12 h sleep deprivation in *cyc⁰¹* flies and *cyc⁰¹* flies pretreated for 3 h at 36 °C. **b**, Cumulative sleep lost or gained during sleep deprivation and subsequent recovery in untreated *cyc⁰¹* flies and preheated flies. **c**, Mortality after 12 h SD in *Hsp83⁰⁸⁴⁴⁵* flies and *Hsp83⁰⁸⁴⁴⁵* flies pretreated for 3 h at 36 °C. **d**, Cumulative sleep lost or gained during sleep deprivation and subsequent recovery in untreated *cyc⁰¹* and *Hsp83⁰⁸⁴⁴⁵*.

*Hsp83*⁰⁸⁴⁴⁵ flies are the only mutants that show both an exaggerated homeostatic response and death after sleep deprivation. Although it is unlikely that these are the only two genes involved in the sleep homeostat, it is worth noting that their mammalian homologues have been shown to interact physically²⁸. Nevertheless, it is possible that the increased homeostatic response and lethality that we observed in *cyc*⁰¹ mutants is due to factors other than *Hsp83*.

Although it is believed that sleep is an essential biological process, its function remains a mystery²⁹. So far, death after chronic total sleep deprivation in the rat provides the best evidence in support of a vital role for sleep³⁰. Our data reinforce these findings and indicate that the vital role of sleep extends beyond mammals; the data also indicate a connection between vulnerability to sleep loss and increased homeostatic drive. Most importantly, the observation that the induction of certain chaperone proteins protects against the lethal effects of sleep loss provides a first hint about the functional targets of sleep and its molecular mechanisms. □

Methods

Source and maintenance of flies

Flies were cultured at 25 °C, 50–60% humidity, 12 h:12 h light:dark cycle, on yeast, dark corn syrup and agar food as described⁸. *per*⁰¹, *yw*; *tim*⁰¹, *Clk*^{trk}*sc*, *cyc*⁰¹; *ry*, *yw*; *cyc*⁰ and *cyc*^{01st} flies were obtained from J.C. Hall (Brandeis University), *per*⁸; *cyc*⁰¹ flies from P. Hardin (University of Houston), and *Df(3L)kto2/TM6B*, *Tb*¹, *Hsp83*⁰⁸⁴⁴⁵, *Hsp83*^{et6a} (recessive lethal) and *Hsp60*^{RA75} (recessive lethal) flies from the Bloomington *Drosophila* Stock Center. Sleep-activity patterns were monitored with the *Drosophila* Activity Monitoring System (Trikinetics) as described previously⁸.

Sleep deprivation

To ensure that flies were awake during the sleep deprivation procedure, we developed a system that coupled the Trikinetics activity monitors with the deprivation apparatus. The Sleep Nullifying Apparatus (SNAP) tilted asymmetrically from –60° to +60° such that the sleeping flies were displaced during the downward movement 10 times per minute. Flies were also deprived of sleep by gentle handling; when an individual fly began sleeping the experimenter would gently tap on the tube. Finally, flies were deprived by rotation as described previously⁸.

Procedure

Female flies 3 d old were individually placed into 65-mm glass tubes in the Trikinetics activity monitoring system under DD. Sleep deprivation was conducted after one full day of baseline, either by the SNAP method or by gentle handling. Flies remained in the Trikinetics monitors during baseline, sleep deprivation and recovery. Cumulative difference plots were calculated for each individual fly first by subtracting the minutes of sleep during deprivation and recovery from the corresponding baseline value and summing the difference score with the preceding hour. A negative slope indicates that sleep is being lost; a positive slope indicates sleep gained and a slope of zero indicates that recovery is complete. Sleep rebound was calculated as a ratio of the amount of sleep recovered divided by that lost, that is [(maximum value when the slope was zero – minimum value)/minimum value]. Statistical significance was assessed for sleep rebound by using a one-way analysis of variance (ANOVA) for genotype. The total numbers of replications and the number of flies of each genotype that were sleep deprived with each method are as follows. For 3 h, gentle handling 5 replications, 80 flies; SNAP 8 replications, 128 flies. For 6 h, gentle handling 4 replications, 64 flies; SNAP 6 replications, 96 flies. For 9 h, gentle handling 1 replication, 16 flies; SNAP 4 replications, 64 flies. For 12 h, gentle handling 2 replications, 32 flies; SNAP 14 replications, 224 flies; and for the rotator 1 replication, 26 flies.

Stress tests

All stress tests were conducted on 3-day-old female flies. During each test, three vials containing 10 flies each were evaluated for each genotype. During each reading, the number of dead flies in a vial was expressed as percentage of the total number of flies. The mean ± s.e.m. of the three vials was calculated for each genotype. A representative example from one of four independent replications for each stress test is shown in Fig. 2. Thus a total of 120 flies were evaluated for each genotype for each stress test. Heat tests were performed at 36 °C in flies maintained on 1% agar, 5% sucrose. Oxidative stress was evaluated in flies maintained at 25 °C in vials with 20 mM paraquat dissolved in 1% agar, 5% sucrose. Flies were starved by placing them in vials with a Kimwipe saturated with water. Desiccation was produced in vials without food or water. Physical stress was evaluated by vortex-mixing flies at high speed for 6 min, after which the flies were allowed 1 min to recover and the number of incapacitated flies was counted. Flies were then vortex-mixed for 2 min and given 1 min to recover, during which time incapacitated flies were again counted. This protocol was repeated until all flies were dead.

QPCR

Total RNA was isolated from fly heads with Trizol (Invitrogen, Carlsbad, California) by following the manufacturer's protocol. Reverse transcription reactions were performed in parallel on DNase-I-digested total RNA as described previously²². Reverse transcription

products were stored at –80 °C until use. PCR reactions to measure levels of artificial transcript were done to confirm uniformity of reverse transcription within sample groups and between samples. Comparable reverse transcription reactions within a sample group were pooled. All reverse transcriptions were performed in quadruplicate. A minimum of three QPCR replications were performed for the sleep deprivation experiments and two for the heat experiments. Values were expressed as a percentage of untreated animals and were evaluated by using a one-way ANOVA for genotype. Dunnett post hoc tests were used to identify values that differed significantly from those of the *cyc*⁰¹ flies.

Received 11 December 2001; accepted 18 March 2002.

- Borbely, A. A. & Achermann, P. Sleep homeostasis and models of sleep regulation. *J. Biol. Rhythms* **14**, 557–568 (1999).
- Mistlberger, R. E., Bergmann, B. M., Waldenar, W. & Rechtschaffen, A. Recovery sleep following sleep deprivation in intact and suprachiasmatic nuclei-lesioned rats. *Sleep* **6**, 217–233 (1983).
- Tobler, I., Borbely, A. A. & Groos, G. The effect of sleep deprivation on sleep in rats with suprachiasmatic lesions. *Neurosci. Lett.* **42**, 49–54 (1983).
- Naylor, E. et al. The circadian clock mutation alters sleep homeostasis in the mouse. *J. Neurosci.* **20**, 8138–8143 (2000).
- Antle, M. C. & Mistlberger, R. E. Circadian clock resetting by sleep deprivation without exercise in the Syrian hamster. *J. Neurosci.* **20**, 9326–9332 (2000).
- Konopka, R. & Benzer, S. Clock mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
- Sehgal, A., Price, J. L., Man, B. & Young, M. W. Loss of circadian behavioural rhythms and per RNA oscillations in the *Drosophila* mutant *timeless*. *Science* **263**, 1603–1606 (1994).
- Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M. A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell* **93**, 791–804 (1998).
- Rutila, J. E. et al. CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* period and timeless. *Cell* **93**, 805–814 (1998).
- Castrillon, D. H. et al. Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single P element mutagenesis. *Genetics* **135**, 489–505 (1993).
- Shaw, P. J., Cirelli, C., Greenspan, R. J. & Tononi, G. Correlates of sleep and waking in *Drosophila melanogaster*. *Science* **287**, 1834–1837 (2000).
- Hendricks, J. C. et al. Rest in *Drosophila* is a sleep-like state. *Neuron* **25**, 129–138 (2000).
- Edgar, D. M., Dement, W. C. & Fuller, C. A. Effect of SCN lesions on sleep in squirrel monkeys: Evidence for opponent processes in sleep–wake regulation. *J. Neurosci.* **13**, 1065–1079 (1993).
- Dijk, D. J. & Czeisler, C. A. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neurosci. Lett.* **166**, 63–68 (1994).
- Homyk, T., Szidony, J. & Suzuki, D. T. Behavioral mutants of *Drosophila melanogaster*. *Mol. Gen. Genet.* **177**, 553–565 (1980).
- Armitage, R., Smith, C., Thompson, S. & Hoffman, R. Sex differences in slow-wave activity in response to sleep deprivation. *Sleep Res. Online* **4**, 33–41 (2001).
- Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. & Winter, J. B. Physiological correlates of prolonged sleep deprivation in rats. *Science* **221**, 182–184 (1983).
- Sapolsky, R. M. Stress hormones: good and bad. *Neurobiol. Dis.* **7**, 540–542 (2000).
- Ashburner, M. Patterns of puffing activity in the salivary gland chromosomes of *Drosophila*. *Chromosoma* **31**, 356–376 (1970).
- Marchler, G. & Wu, C. Modulation of *Drosophila* heat shock transcription factor activity by the molecular chaperone DRO1. *EMBO J.* **20**, 499–509 (2001).
- Ornelles, D. A. & Penman, S. Prompt heat-shock and heat-shifted proteins associated with the nuclear matrix–intermediate filament scaffold in *Drosophila melanogaster* cells. *J. Cell Sci.* **95**, 393–404 (1990).
- Cirelli, C. & Tononi, G. Gene expression in the brain across the sleep–waking cycle. *Brain Res.* **885**, 303–321 (2000).
- Yoshida, E. N., Benkel, B. F., Fong, Y. & Hickey, D. A. Sequence and phylogenetic analysis of the SNF4/AMPK gamma subunit gene from *Drosophila melanogaster*. *Genome* **42**, 1077–1087 (1999).
- Ma, E. & Haddad, G. G. Isolation and characterization of the hypoxia-inducible factor 1β in *Drosophila melanogaster*. *Brain Res. Mol. Brain Res.* **73**, 11–16 (1999).
- Cornelius, G. & Engel, M. Stress causes induction of MAP kinase-specific phosphatase and rapid repression of MAP kinase activity in *Drosophila*. *Cell Signal.* **7**, 611–615 (1995).
- Ekgren, S. et al. A humoral stress response in *Drosophila*. *Curr. Biol.* **11**, 714–718 (2001).
- Perezgasga, L., Segovia, L. & Zurita, M. Molecular characterization of the 5' control region and of two lethal alleles affecting the *hsp60* gene in *Drosophila melanogaster*. *FEBS Lett.* **456**, 269–273 (1999).
- Hogensch, J. B. et al. Characterization of a subset of the basic-helix–loop–helix–PAS superfamily that interacts with components of the dioxin signalling pathway. *J. Biol. Chem.* **272**, 8581–8593 (1997).
- Rechtschaffen, A. Current perspectives on the function of sleep. *Perspect. Biol. Med.* **41**, 359–390 (1998).
- Rechtschaffen, A., Bergmann, B. M., Everson, C. A., Kushida, C. A. & Gilliland, M. A. Sleep deprivation in the rat: X. Integration and discussion of the findings. *Sleep* **12**, 68–87 (1989).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank H. Dierick, R. Andretic and K. Long for helpful comments on the manuscript, and J. Hendricks for sharing data before publication. This work was performed at The Neurosciences Institute, which is supported by the Neurosciences Research Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

A common mechanism of action for three mood-stabilizing drugs

Robin S. B. Williams*, Lili Cheng†, Anne W. Mudge†‡
& Adrian J. Harwood*‡

* Intracellular Signalling and † Cellular Neurobiology Groups, MRC Laboratory for Molecular Cell Biology, and Department of Biology, University College London, Gower St, London WC1E 6BT, UK

‡ These authors contributed equally to this work

Lithium, carbamazepine and valproic acid are effective mood-stabilizing treatments for bipolar affective disorder. The molecular mechanisms underlying the actions of these drugs and the illness itself are unknown. Berridge and colleagues¹ suggested that inositol depletion may be the way that lithium works in bipolar affective disorder, but others have suggested that glycogen synthase kinase^{2,3} (GSK3) may be the relevant target. The action of valproic acid has been linked to both inositol depletion^{4,5} and to inhibition of histone deacetylase⁶ (HDAC). We show here that all three drugs inhibit the collapse of sensory neuron growth cones and increase growth cone area. These effects do not depend on GSK3 or HDAC inhibition. Inositol, however, reverses the effects of the drugs on growth cones, thus implicating inositol depletion in their action. Moreover, the development of *Dictyostelium* is sensitive to lithium⁷ and to valproic acid, but resistance to both is conferred by deletion of the gene that codes for prolyl oligopeptidase, which also regulates inositol metabolism. Inhibitors of prolyl oligopeptidase reverse the effects of all three drugs on sensory neuron growth cone area and collapse. These results suggest a molecular basis for both

bipolar affective disorder and its treatment.

In addition to their use as mood stabilizers, carbamazepine (CBZ) and valproic acid (VPA) are also used to treat epilepsy, and VPA⁶ and lithium² are teratogenic. Their different effects have made it difficult to establish whether the three drugs act on similar cell and molecular targets to stabilize mood in bipolar affective disorder. We reasoned that if it were possible to identify common cellular effects of these structurally diverse mood stabilizers, this might suggest which molecular targets are relevant for their effects on mood and hence the underlying defect in bipolar affective disorder. The cellular effects of lithium have been studied extensively using cultured mammalian neurons as well as lower organisms, but there have been few comparable studies with CBZ and VPA. We therefore reinvestigated the effects of lithium on cultured neurons and compared them with those of CBZ and VPA.

We cultured explants of sensory neurons from newborn rat dorsal root ganglia and tested the drugs at concentrations similar to their recommended therapeutic plasma levels⁸. Consistent with previous studies^{3,9,10}, lithium treatment resulted in unbundled microtubules in axons, which frequently adopted a zig-zag trajectory (Fig. 1a, left panel). About 2% of growth cones became very large, with microtubules extending abnormally into the growth cones (Fig. 1a, right panel). These effects on microtubule structure were not observed with CBZ or VPA treatment. Lithium also caused a large increase in the number of collateral axon branches that formed proximal to the growth cones (Fig. 1a, b), whereas CBZ had no effect and VPA decreased the number of such branches (Fig. 1b). When lithium and VPA were used together, axons had the appearance of those treated with VPA alone (Fig. 1b).

The HDAC inhibitor trichostatin A (TSA) mimics the teratogenic effects of VPA⁶. TSA also had the same effect as VPA in decreasing the number of collateral axon branches, suggesting that inhibition

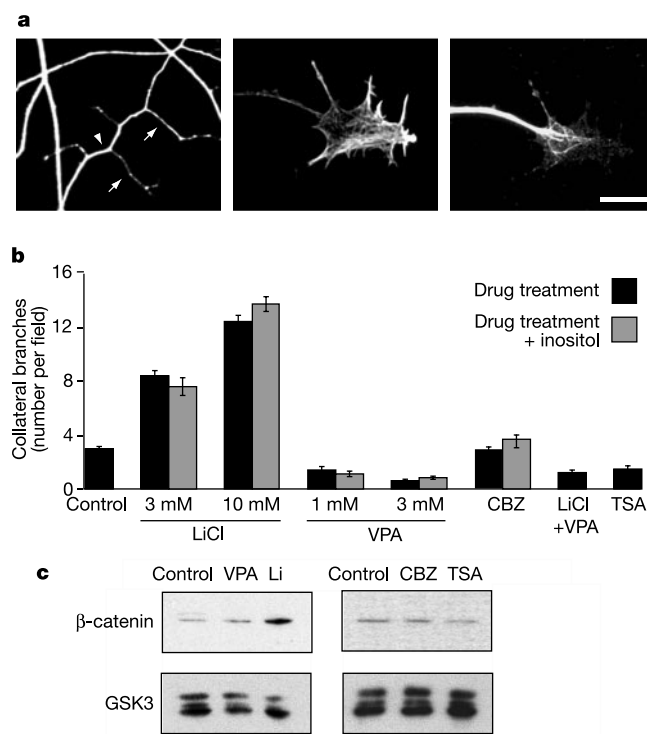


Figure 1 Structural changes in sensory neuron axons following treatment with mood-stabilizing drugs. **a**, Micrograph (left panel) showing axons with a zig-zag trajectory (arrowhead) and collateral branches (arrows) in lithium-treated cultures. Micrographs of an abnormal growth cone double-labelled with phalloidin (middle panel) and anti-acetylated tubulin (right panel), showing coiled microtubules in the growth cone of a

lithium-treated axon. Scale bar, 10 μ m. **b**, Histograms showing the average number of collateral branches per field after treatment with drugs as indicated (mean $\pm$ s.e.m., $n = 60$ fields). **c**, Western blot analysis of β -catenin and GSK3 levels in sensory neurons after drug treatment as indicated. VPA, valproic acid; CBZ, carbamazepine; TSA, trichostatin A.

of axonal branching by VPA depends on the inhibition of HDAC (Fig. 1b). Lithium is thought to alter microtubule and axonal structure by inhibition of GSK3 (ref. 3), and deletion of the genes that code for the microtubule-associated proteins tau or MAP1B—both substrates of GSK3—leads to abnormalities in growth cones similar to those seen with lithium treatment¹¹. Lithium, but not CBZ or VPA, inhibits the activity of both the α and β isoforms of GSK3 (ref. 2) (W. J. Ryves and A. J. H., unpublished data). β -catenin is regulated in most cells by GSK3 (refs 12, 13) and by HDAC in a neuroblastoma cell line⁶. Lithium, but not CBZ or VPA, increased β -catenin in our dorsal root ganglia cultures (Fig. 1b). Thus, although both GSK3 and HDAC influence axon growth, their inhibition changes axon morphology in different ways, and they are unlikely to be a common target for lithium, CBZ and VPA.

All three drugs, however, had similar effects on the dynamic behaviour of sensory neuron growth cones. When untreated cultures were viewed by time-lapse video microscopy, the growth cones had pronounced cycles of complete collapse followed by active spreading. Each of the drugs reduced the frequency of collapse and the growth cones were enlarged following each treatment compared with controls (L.C. and A.W.M., unpublished data). The increase in growth cone size may be a consequence of the effect of inhibiting collapse. We quantified these effects by analysing the frequency distribution of collapsed and spread growth cones using fixed and labelled cultures (Fig. 2a, b). In untreated cultures, 27% of growth cones were collapsed (first bar in histograms), whereas CBZ and VPA reduced this proportion five fold, to about 5%, while lithium decreased it 2–3-fold. Lithium, CBZ and VPA also increased the average spread area of growth cones by 56%, 63% and 81%, respectively. These effects on growth cones have not been reported in previous studies with lithium, perhaps because we unmasked the effects by using serum-free medium, where we find that collapse is more pronounced than in serum-containing medium. To our knowledge, this is the first report of a common effect of all three drugs on neurons.

Lithium inhibits inositol monophosphatase (IMPase) and inositol polyphosphatase, thereby leading to a decreased inositol-1,4,5-trisphosphate (InsP_3) response¹. This inhibition prevents both recycling of inositol from inositol phosphates and *de novo* inositol synthesis, and can be overcome by addition of extracellular inositol. The effects of the three drugs on growth cone collapse and area were abolished by addition of inositol, implicating inositol phosphate signalling in this common response (Fig. 2b, right panels). In contrast, addition of inositol had no effect on either the lithium-induced increase in collateral axon branching and giant growth cones, or the VPA-induced decrease in collateral branching (Fig. 1b); this argues against a role for inositol depletion in any of these processes. TSA had no effect on the growth cones, and we showed above that neither CBZ nor VPA affects GSK3; thus, neither HDAC nor GSK3 are likely to be involved in mediating the drug effects on the dynamic behaviour of growth cones.

In all our studies, we observed significant effects at drug concentrations close to the recommended therapeutic plasma levels⁸ (0.6–1.0 mM lithium, 0.3–0.6 mM VPA and 20–50 μM CBZ). The levels in clinical use, however, are limited by side effects, and they do not necessarily elicit maximum mood-stabilizing effects. Lithium and VPA are used at remarkably high concentrations: VPA, however, binds to protein and membranes, and the free concentration may be much lower. In the case of lithium, it is notable that the inhibition constant, K_i , for IMPase and GSK3 is 0.8 mM and 2 mM, respectively. Although our studies used growth cones that are characteristic of developing neurons, we think it likely that the effects we describe on inositol depletion will also apply to neurons in the mature brain.

Dictyostelium offers a genetic route to the analysis of drug action. During early development, cells aggregate by chemotaxis to form a multicellular mass. Aggregation is defective when InsP_3 signalling is disrupted by lithium, whereas the effects of lithium on later

development are due to GSK3 inhibition. Late development is also sensitive to VPA¹⁴, but the mechanism is unknown. We found that VPA reduced basal InsP_3 concentration (Fig. 3a). Transcription of the inositol-1-phosphate synthase (*ino1*) gene in yeast is repressed by inositol, and so is an *in vivo* indicator of inositol concentration⁵. As shown in Fig. 3b, 10 mM lithium caused a twofold increase in *ino1* expression, whereas 1 mM VPA produced

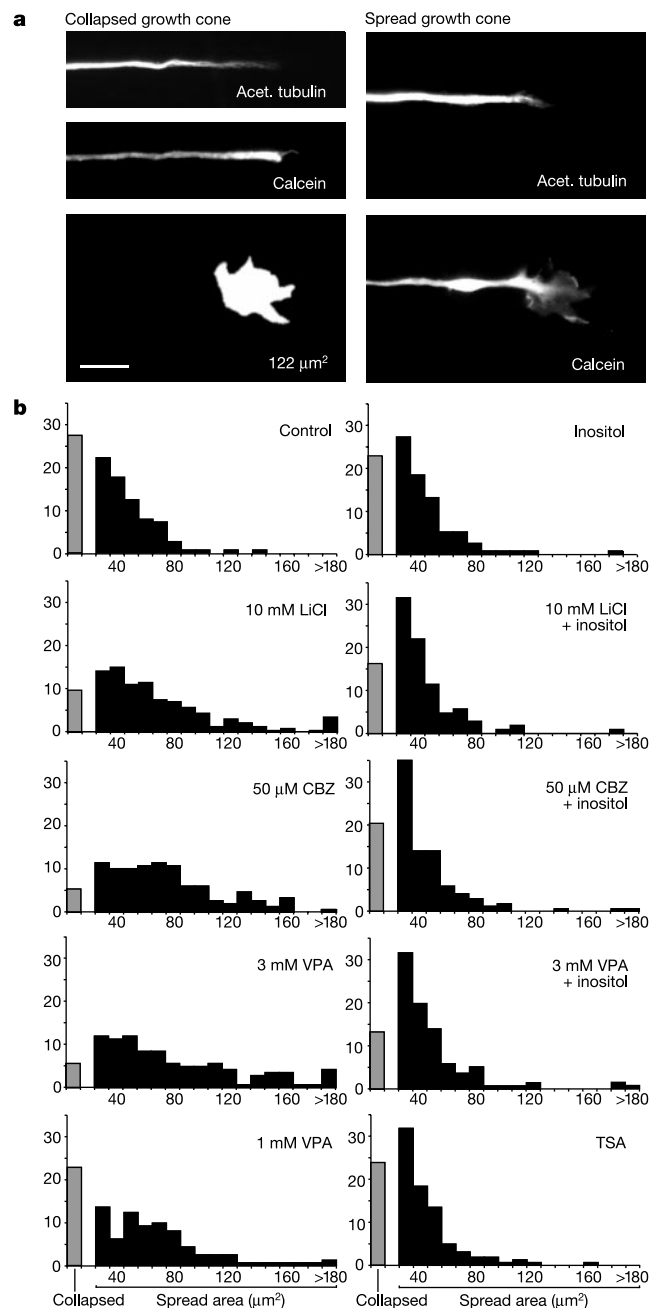


Figure 2 Inositol reverses the effects of mood-stabilizing drugs on growth cone collapse and spread area. **a**, Collapsed growth cones (top left panels) and spread growth cones (right panels) showing stable microtubules (acetylated tubulin) and cytoplasm (calcein). The area of this growth cone was calculated as $122 \mu\text{m}^2$ as shown (bottom left panel). Scale bar, $10 \mu\text{m}$. **b**, Histograms show the frequency distribution of collapsed growth cones (grey first bar) and the spread area of growth cones plotted in increments of $10 \mu\text{m}^2$ (black bars) expressed as a percentage of the total. The drug treatment for left-hand panels is indicated on each plot. The right-hand panels are from cultures treated with drugs in the presence of 1 mM inositol, or cultures treated with TSA alone as indicated. Data were pooled from up to 5 different platings; the total number of growth cones scored for each treatment was about 150.

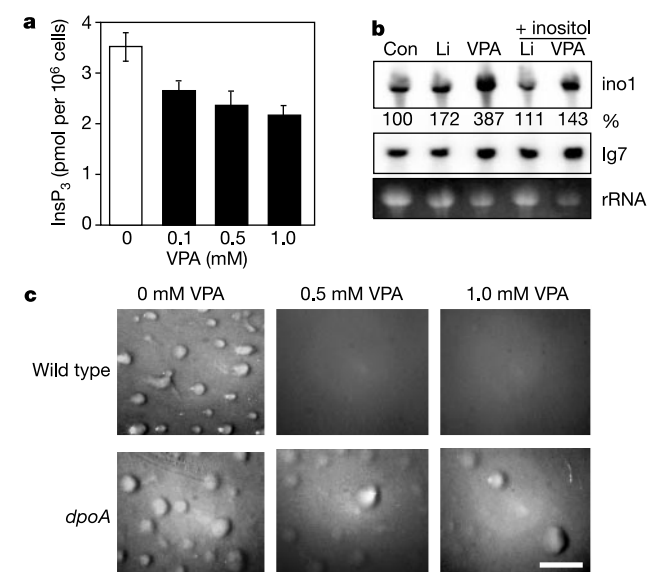


Figure 3 VPA effects on *Dictyostelium* aggregation are due to modulation of InsP_3 signalling. **a**, Basal InsP_3 levels were measured with and without treatment with VPA. **b**, Expression of the inositol-1-phosphate synthase gene was assessed by northern analysis. Cells were treated with 10 mM lithium, 1 mM VPA, in the absence or presence of 10 mM inositol. Increased *ino1* expression is shown relative to Ig7 (loading control). Ethidium stained rRNA is also shown. **c**, Aggregation of wild-type *Dictyostelium* cells and a mutant lacking the gene that codes for prolyl oligopeptidase (*dpoA*) in 0, 0.5 and 1 mM VPA.

a fourfold increase: both treatments were reversed by inositol addition. This indicates that both drugs reduce the *in vivo* concentration of inositol.

We previously used insertional mutagenesis to generate lithium-resistant *Dictyostelium* mutants⁷. Loss of the gene *dpoA*, which encodes a homologue of the enzyme prolyl oligopeptidase, confers lithium resistance⁷ and is associated with elevated basal levels of InsP_3 . As shown in Fig. 3c, loss of *dpoA* also conferred cross-

resistance to VPA, strongly suggesting that InsP_3 is a major target of VPA in *Dictyostelium*, and that elevation of basal InsP_3 can counteract the VPA effect on aggregation. These direct biochemical measurements and genetic interactions support our conclusion that VPA can target inositol phosphate metabolism in sensory neurons.

Prolyl oligopeptidase is a cytoplasmic protein in both mammals and *Dictyostelium*. It can cleave short oligopeptides at prolyl bonds, but its function is not known. Plasma concentrations of this protein are abnormal in people with affective disorders^{15,16}—they are elevated in mania and decreased in depression—although the significance of the finding is unclear. We therefore tested whether there is a link between prolyl oligopeptidase and the action of the mood-stabilizing drugs on neurons. We used two specific inhibitors of prolyl oligopeptidase activity^{17,18}, and found that both abolished the effects of lithium, CBZ and VPA on growth cone collapse and area. The frequency distributions in Fig. 4 show that there is no significant difference between control growth cones and those treated with drugs in the presence of the inhibitors. In all cases, the number of collapsed growth cones was 26% compared with 27% for untreated controls. The spread area of growth cones in the presence of drugs plus inhibitors was $42 \mu\text{m}^2$ compared with untreated controls, which was $43 \mu\text{m}^2$. This result suggests that prolyl oligopeptidase may be involved in regulating inositol metabolism in mammalian cells.

Our findings support the hypothesis¹ that the therapeutic target of lithium in the treatment of bipolar affective disorder depends on inositol depletion, and further extends the hypothesis to CBZ and VPA. The brain is highly sensitive to inositol depletion as the blood-brain barrier limits the availability of plasma inositol, making it dependent on inositol recycling and synthesis. The association between prolyl oligopeptidase activity and affective disorders is intriguing and should be explored further, particularly in view of our finding that inhibition of this enzyme mimics the action of added inositol in reversing the drug effects. Finally, these results suggest that the development of both clinical diagnostic tests and new therapies for bipolar affective disorder should focus on inositol phosphate metabolism. □

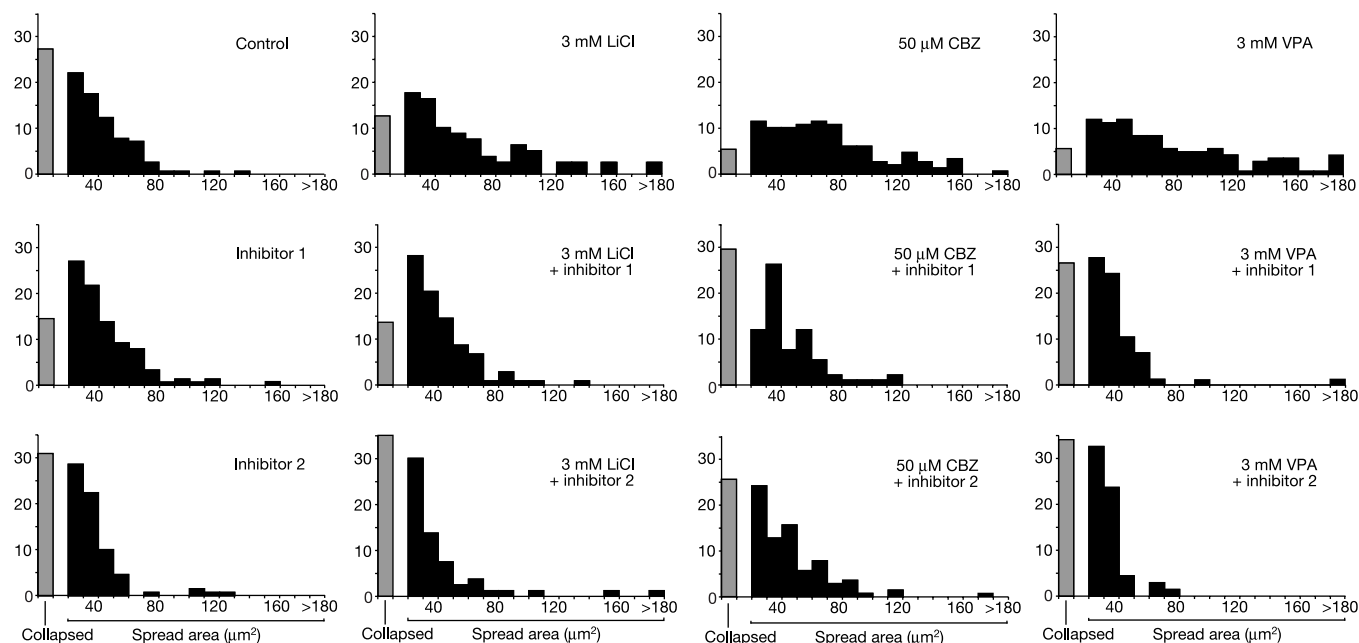


Figure 4 Inhibitors of prolyl oligopeptidase enzyme activity block the action of mood-stabilizing drugs on growth cone collapse and spread area. Histograms show the frequency distribution of collapsed growth cones (grey first bar) and the spread area of growth cones plotted in increments of $10 \mu\text{m}^2$ (black bars) expressed as a percentage of

the total. The top row shows the effects of the drugs. The second and third rows show the effects of the drugs when they were added in the presence of two structurally unrelated prolyl oligopeptidase inhibitors, which are each specific for the prolyl bond. Inhibitors 1 and 2 are defined in Methods.

Methods

Neuron explant cultures

Dorsal root ganglia from newborn rats were cultured as explants on laminin-coated coverslips in serum-free medium¹⁹ with nerve growth factor (NGF) at 50 µg ml⁻¹ (mouse 7S form, Alomone Labs). Cytosine-arabinoxanthine (ara-C, 10 µM) was added after 1 d in culture to kill non-neuronal cells. The drugs to be tested—lithium chloride, 2-propylpentanoic acid (valproic acid, VPA), carbamazepine (CBZ), trichostatin A (TSA), and myo-inositol (all from Sigma)—were also added after 1 d and explants were cultured for a further 2 d before fixation or collection. Trichostatin A was added at 25 µM. Prolyl oligopeptidase inhibitors Z-pro-pro-aldehyde-dimethyl acetal (inhibitor 1, Bachem) and BOC-Glu (NHO-Bz)-Pyr (inhibitor 2, Calbiochem) were added at 133 µM and 24 µM, respectively. Inositol (in the form of myo-inositol) was added at 1 mM.

Immunocytochemistry

Cells were loaded with the fluorescent dye Calcein (Molecular Probes) for 20 min and then fixed with 4% paraformaldehyde and 0.2% glutaraldehyde in cytoskeletal buffer²⁰ containing 10 mM MES buffer at pH 6.1 with 138 mM KCl, 3 mM MgCl₂, and 2 mM EGTA buffer; aldehydes were quenched with sodium borohydride. Cells were then labelled with monoclonal antibody to acetylated tubulin (Sigma, clone 6-11B-1) followed by anti-mouse immunoglobulin, then biotin-streptavidin conjugated to Texas red. In some cases, cells were labelled with fluorescent phalloidin (Alexa⁵⁹⁴-phalloidin) followed by anti-acetylated tubulin visualized with fluorescein.

Neuron data collection

We scored random fields at the perimeter of the axonal halo that grew out from the explanted cell bodies. Collateral axon branches were counted using acetylated tubulin as the label, and growth cones were scored using the cytoplasmic dye calcein. The growth cone perimeter was traced with a light pen, and the area was computed using OpenLab software. Collapsed growth cones with no calcein-labelled lamellae extending beyond the microtubules were assigned an arbitrary value of 1 for computational purposes. Experiments were done with triplicate coverslips for each treatment, and data were pooled from up to five different platings to generate the histograms shown.

Western blotting

Samples were separated by 10% SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose membrane (Amersham Life Sciences). Western blots were probed with anti-β-catenin²¹ (C19220, Transduction Laboratories) or anti-GSK3 (4G-1E, Upstate Biotechnology) primary antibodies. Blots were then labelled with anti-mouse IgG-HRP antibody (Vector Laboratories) and visualized using SuperSignal West Pico Luminol/Enhancer solution (Pierce) according to the supplier's instructions.

Dictyostelium cell culture and development

Wild-type *Dictyostelium discoideum* cells (Ax2) or *dpoA* mutant cells⁷ were grown at 22 °C in shaking culture in HL5 medium to 3 × 10⁶ cells ml⁻¹ or in association with *Klebsiella aerogenes*. Effects of drug treatment on aggregation was analysed by plating cells at 2.5 × 10⁶ per 47 mm nitrocellulose filter (Millipore) soaked in KK₂ (16.5 mM KH₂PO₄, 3.8 mM K₂HPO₄ pH 6.2) containing lithium or VPA (buffered to pH 7.0) as indicated. Development was observed after 8 h.

InsP₃ assays

Wild-type cells were grown in shaking suspension for 20 h at 2 × 10⁶ cells ml⁻¹ containing defined concentrations of VPA. Cells were re-suspended at 5 × 10⁷ cells ml⁻¹ and aerated for 10 min. Cells (200 µl) were lysed and the InsP₃ concentrations measured by an isotope dilution binding assay (Amersham Pharmacia Biotech).

Northern analysis

Cells were treated with VPA or LiCl as described for InsP₃ assays, and RNA was prepared as previously described²². The *ino1* cDNA (inositol-1-phosphate synthase, clone SLB 678), provided by the *Dictyostelium* cDNA project in Japan, was labelled by random priming, and hybridized at 40 °C in hybridization solution containing 42% formamide²³.

Received 26 January; accepted 5 March 2002.

- Berridge, M. J., Downes, C. P. & Hanley, M. R. Neural and developmental actions of lithium: a unifying hypothesis. *Cell* **59**, 411–419 (1989).
- Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc. Natl Acad. Sci. USA* **93**, 8455–8459 (1996).
- Lucas, F. R. & Salinas, P. C. WNT-7a induces axonal remodeling and increases synapsin I levels in cerebellar neurons. *Dev. Biol.* **192**, 31–44 (1997).
- O'Donnell, T. et al. Chronic lithium and sodium valproate both decrease the concentration of myo-inositol and increase the concentration of inositol monophosphates in rat brain. *Brain Res.* **880**, 84–91 (2000).
- Vaden, D. L., Ding, D., Peterson, B. & Greenberg, M. L. Lithium and valproate decrease inositol mass and increase expression of the yeast *ino1* and *ino2* genes for inositol biosynthesis. *J. Biol. Chem.* **276**, 15466–15471 (2001).
- Phiel, C. J. et al. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J. Biol. Chem.* **276**, 36734–36741 (2001).
- Williams, R. S., Eames, M., Ryves, W. J., Viggars, J. & Harwood, A. J. Loss of a prolyl oligopeptidase confers resistance to lithium by elevation of inositol (1,4,5) trisphosphate. *EMBO J.* **18**, 2734–2745 (1999).
- Kerwin, R. *The Bethlem and Maudsley NHS Trust Prescribing Guidelines* (Duntiz, London, 1999).
- Hall, A. C., Lucas, F. R. & Salinas, P. C. Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling. *Cell* **100**, 525–535 (2000).
- Goold, R. G., Owen, R. & Gordon-Weeks, P. R. Glycogen synthase kinase 3β phosphorylation of microtubule-associated protein 1B regulates the stability of microtubules in growth cones. *J. Cell Sci.*

112, 3373–3384 (1999).

- Takei, Y., Teng, J., Harada, A. & Hirokawa, N. Defects in axonal elongation and neuronal migration in mice with disrupted tau and map1b genes. *J. Cell. Biol.* **150**, 989–1000 (2000).
- Stambolic, V., Ruel, L. & Woodgett, J. R. Lithium inhibits glycogen synthase kinase-3 activity and mimics wingless signalling in intact cells. *Curr. Biol.* **6**, 1664–1668 (1996).
- Yost, C. et al. The axis-inducing activity, stability, and subcellular distribution of beta-catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
- Tillner, J., Nau, H., Winckler, T. & Dingermann, T. Evaluation of the teratogenic potential of valproic acid analogues in transgenic *Dictyostelium discoideum* strains. *Toxicol. Vitro* **12**, 463–469 (1998).
- Maes, M. et al. Lower serum prolyl endopeptidase activity in major depression: further evidence that peptidases play a role in the pathophysiology of depression. *Biol. Psychiatry* **35**, 545–552 (1994).
- Maes, M. et al. Alterations in plasma prolyl endopeptidase activity in depression, mania, and schizophrenia: effects of antidepressants, mood stabilizers, and antipsychotic drugs. *Psychiatry Res.* **58**, 217–225 (1995).
- Augustyns, K. et al. Synthesis of peptidyl acetals as inhibitors or prolyl endopeptidase. *Bioorg. Medicinal Chem. Lett.* **5**, 1265–1270 (1995).
- Demuth, H. U. et al. Design of (omega-N-(O-acyl)hydroxy amid) aminodicarboxylic acid pyrrolidides as potent inhibitors of proline-specific peptidases. *FEBS Lett.* **320**, 23–27 (1993).
- Cheng, L. & Mudge, A. W. Cultured Schwann cells constitutively express the myelin protein P0. *Neuron* **16**, 309–319 (1996).
- Cramer, L. P., Siebert, M. & Mitchison, T. J. Identification of novel graded polarity actin filament bundles in locomoting heart fibroblasts: implications for the generation of motile force. *J. Cell Biol.* **136**, 1287–1305 (1997).
- Kypta, R. M., Su, H. & Reichardt, L. F. Association between a transmembrane protein tyrosine phosphatase and the cadherin-catenin complex. *J. Cell Biol.* **134**, 1519–1529 (1996).
- Harwood, A. J., Plyte, S. E., Woodgett, J., Strutt, H. & Kay, R. R. Glycogen synthase kinase 3 regulates cell fate in *Dictyostelium*. *Cell* **80**, 139–148 (1995).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Press, 1989).

Acknowledgements

This study was an equal collaboration between the Harwood and Mudge research groups, and the paper was co-written by A.W.M. and A.J.H. The neuron experiments were carried out by L.C., and *Dictyostelium* experiments by R.S.B.W. and A.J.H. The neuron analysis was designed by A.W.M., and carried out by R.S.B.W. and A.W.M. The work was supported by both MRC (L.C., A.W.M.) and Wellcome Trust funding (A.J.H. and R.S.B.W.). We thank L. Cramer for discussions. We also thank M. Shipman and B. Mudge for help with microscopy and graphics, respectively.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.W.M. (e-mail: a.mudge@ucl.ac.uk).

The Wnt/calcium pathway activates NF-AT and promotes ventral cell fate in *Xenopus* embryos

Takeo Saneyoshi*, Shoen Kume*†‡, Yoshiharu Amasaki§ & Katsuhiko Mikoshiba*†||¶

* Department of Molecular Neurobiology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
† Mikoshiba Calciosignal Net Project, ERATO and || Calcium Oscillation Project, ICORP, Japan Science and Technology Corporation (JST), 3-14-4 Shirokanedai, Minato-ku, Tokyo 108-0071, Japan

§ Department of Molecular and Developmental Biology, Institute of Medical Science, The University of Tokyo, CREST, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

¶ RIKEN Brain Research Institute, 2-1 Hirosawa, Wako, Saitama 351-01, Japan
‡ Present address: Division of Stem Cell Biology, Department of Regeneration Medicine, Institute of Molecular Embryology and Genetics, Kumamoto University, Kuhonji 4-24-1, Kumamoto 862-0976, Japan

It is thought that inositol-1,4,5-trisphosphate (Ins(1,4,5)P₃)-Ca²⁺ signalling has a function in dorsoventral axis formation in *Xenopus* embryos^{1–3}; however, the immediate target of free Ca²⁺ is unclear. The secreted Wnt protein family comprises two functional groups, the canonical Wnt and Wnt/Ca²⁺ pathways⁴. The Wnt/Ca²⁺ pathway interferes with the canonical Wnt path-

way⁵, but the underlying molecular mechanism is poorly understood. Here, we cloned the complementary DNA coding for the *Xenopus* homologue of nuclear factor of activated T cells (XNF-AT). A gain-of-function, calcineurin-independent active XNF-AT mutation (CA XNF-AT) inhibited anterior development of the primary axis, as well as Xwnt-8-induced ectopic dorsal axis development in embryos. A loss-of-function, dominant negative XNF-AT mutation (DN XNF-AT) induced ectopic dorsal axis formation and expression of the canonical Wnt signalling target molecules *siamoi*s and *Xnr3* (ref. 4). Xwnt-5A induced translocation of XNF-AT from the cytosol to the nucleus. These data indicate that XNF-AT functions as a downstream target of the Wnt/Ca²⁺ and Ins(1,4,5)P₃-Ca²⁺ pathways, and has an essential role in mediating ventral signals in the *Xenopus* embryo through suppression of the canonical Wnt pathway.

Gain of function of Ins(1,4,5)P₃-Ca²⁺ signalling on the dorsal side of the embryo leads to a dorso-anterior structure deficiency³, whereas loss of function on the ventral side induces a partial ectopic dorsal axis². These findings suggest that the Ins(1,4,5)P₃-Ca²⁺ signalling pathway mediates ventral signals. One possible target of Ins(1,4,5)P₃-Ca²⁺ signalling is the Ca²⁺/calmodulin (CaM)-dependent protein phosphatase, calcineurin, and the transcription factor—positioned further downstream—NF-AT⁶. To study downstream signalling of the Ins(1,4,5)P₃-Ca²⁺ pathway with respect to dorsoventral axis formation through a ventralizing signal, we cloned a *Xenopus* homologue of NF-AT (XNF-AT). Protein alignment showed that XNF-AT was 49.2%, 49.1%, 64.1% and 39.5% identical at the amino-acid level to human NF-ATc1, NF-ATc2, NF-ATc3 and NF-ATc4, respectively (see Supplementary Information). XNF-AT protein is expressed as a dephosphorylated form throughout early stages of cleavage. At the start of mesoderm induction in *Xenopus* (between 16- to 32-cell stages^{7,8}), we observed a marked, transient switch to a phosphorylated form of XNF-AT (Fig. 1a). After the gastrula stage, the phosphorylated form of XNF-AT gradually increased, an event that correlated well with decreased Ins(1,4,5)P₃ receptor protein expression (Fig. 1a). XNF-AT, which was dephosphorylated by calcium ionophore A23187 stimulation, was cancelled by FK506, a calcineurin blocker; thereby indicating that dephosphorylation of XNF-AT is mediated by the action of calcineurin (Fig. 1b). Furthermore, XNF-AT phosphorylation was observed when embryos were depleted of calcium by BAPTA injection into embryos (T.S., unpublished data). These data suggest that the dephosphorylated form of XNF-AT at the 16–32-cell stage is in a default state, and that transient increase with the phosphorylated form of XNF-AT at the 16–32-cell stage might be due to a transient inactivation of XNF-AT activity, as mediated by calcineurin inactivation.

To better understand the function of XNF-AT, we generated a dominant negative mutant XNF-AT (DN XNF-AT) based on the homology to murine NF-AT^{9,10}, and a calcineurin-independent, active mutant XNF-AT (CA XNF-AT) (Fig. 1c). Wild-type XNF-AT (WT XNF-AT) was observed to translocate from the cytosol to the nucleus in COS7 cells on A23187 stimulation, whereas the constitutively active mutant CA XNF-AT was retained in the nucleus and was independent of A23187 stimulation (Fig. 1d). Transcriptional activities of XNF-AT mutants were assayed by NF-AT-dependent activation of a murine interleukin (IL)-2 promoter¹¹ in *Xenopus* embryos. The CA XNF-AT mutant showed transcriptional activity independent of calcineurin (that is, CA XNF-AT requires only AP1), whereas WT XNF-AT requires both constitutively active *Xenopus* calcineurin (XΔCnA)¹² and AP1 to activate the IL-2 promoter (Fig. 1e, left). DN XNF-AT showed dose-dependent inhibition of WT XNF-AT transcriptional activity (Fig. 1e, right), thus confirming its function as a dominant negative mutant.

Injection of CA XNF-AT RNA led to ventralized embryos, whereas injection of WT XNF-AT RNA had no effect (Fig. 2a). The effect of CA XNF-AT RNA occurred through inhibition of

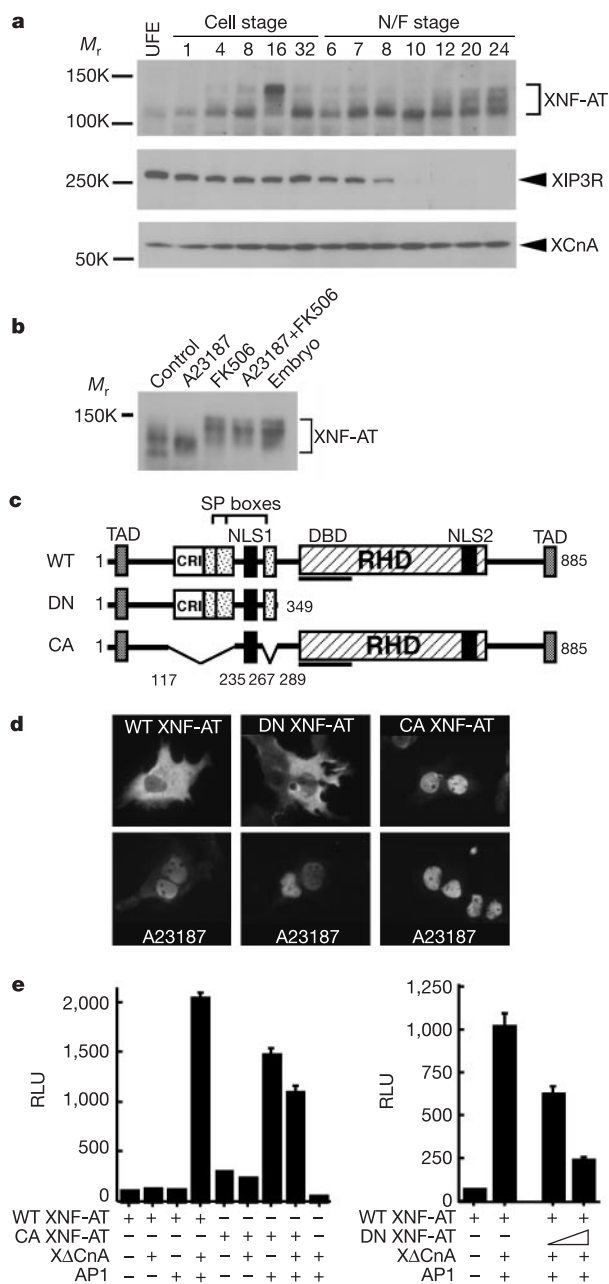


Figure 1 XNF-AT expression and regulation by calcineurin. **a**, Western blot analysis of the expression of XNF-AT (top), Ins(1,4,5)P₃ receptor (XIP3R; middle) and XCnA (bottom) during embryogenesis. N/F, Nieukoop and Faber. UFE, unfertilized egg. **b**, Effect of the calcineurin blocker FK506 (10 μM) and A23187 (1 μM) on XNF-AT protein in the animal cap. **c**, Schematic representation of XNF-AT mutants. WT, wild-type XNF-AT; DN, dominant negative XNF-AT; CA, constitutively active XNF-AT. TAD, transactivation domain; DBD, DNA binding domain; RHD, Rel homology domain. **d**, Subcellular localization of WT XNF-AT, DN XNF-AT and CA XNF-AT in response to A23187 (0.5 μM) in COS7 cells. **e**, CA XNF-AT shows calcineurin-independent transcriptional activity (left). DN XNF-AT inhibited XNF-AT-mediated transcriptional activation in a dose-dependent manner (right). XΔCnA, constitutively active calcineurin.

morphogenetic movements, as CA XNF-AT blocked activin A- or basic fibroblast growth factor (bFGF)-induced animal cap elongation (Fig. 2a) without interference with mesoderm induction, as indicated by the expression of the pan-mesodermal marker *Xbra*, and organizer markers *chordin* and *gooseoid* (Fig. 2b). Injection of DN XNF-AT RNA dose-dependently induced a complete secondary dorsal axis (Fig. 2c, DN XNF-AT; see also Fig. 2d), which was rescued by co-injection of WT XNF-AT RNA. Thus,

induction of the dorsal axis by DN XNF-AT occurs through specific interference with the function of endogenous XNF-AT (Fig. 2e). Taken together, these results suggest that XNF-AT is an important regulator of ventral cell fate.

A candidate for the upstream ventralizing signal that may activate $\text{Ins}(1,4,5)\text{P}_3\text{-Ca}^{2+}$ signalling is the Wnt/ Ca^{2+} pathway (for example, Wnt-4, -5A and -11)⁴. It has been reported that activation of the Wnt/ Ca^{2+} pathway leads to an increase in intracellular Ca^{2+} in zebrafish¹³ and inhibition of activin A-induced morphogenetic movement, without blocking mesoderm induction¹⁴. Furthermore, loss-of-function assays for the Wnt/ Ca^{2+} pathway showed that a dominant negative mutant of Xwnt-11 induced axis duplication¹⁵, and that antisense Wnt-5A mimicked Wnt-1 activity in cultured cells¹⁶. The phenotypes we observed here with XNF-AT mutants suggest that XNF-AT is the downstream target of the Wnt/ Ca^{2+} pathway (Fig. 2). To address this question, we examined subcellular

localization and embryonic effects of XNF-AT in response to the Wnt/ Ca^{2+} pathway. Myc-tagged XNF-AT localized mainly in the cytoplasm in a resting state (Fig. 3a). Translocation of XNF-AT to the nucleus occurred on co-expression with Xwnt-5A and its receptor rat frizzled 2 (Rfz2), and which was also observed with X ΔCnA^{12} (Fig. 3a). Embryos injected with Xwnt-5A exhibited the dorsally bent axis shown in Fig. 3b (Xwnt-5A phenotype, 98%, $n = 71$). This dorsal defect was similar to events seen in embryos injected with CA XNF-AT (dorso-anterior index, DAI¹⁷, 0–3, 66.7%, $n = 57$) (Fig. 3b, CA XNF-AT), and was enhanced on co-injection of WT XNF-AT with Xwnt-5A RNA (Xwnt-5A phenotype, 23.7%, DAI 0–3, 66.1%, $n = 59$). These results indicate that stimulation of the Wnt/ Ca^{2+} pathway activates calcineurin and leads to translocation of XNF-AT from the cytosol to the nucleus.

As CA XNF-AT has a phenotype similar to that of Xwnt-5A¹⁴ (Figs 2a, b and 3c), XNF-AT signalling may negatively regulate the canonical Wnt pathway⁵. To test this, we determined whether CA XNF-AT would inhibit the Xwnt-8-induced ectopic axis formation. The frequency of complete axis duplication by Xwnt-8 (73.7%, $n = 205$) was reduced to 14.1% by co-injection with CA XNF-AT ($n = 78$), whereas WT XNF-AT had no effects ($n = 103$). Next, we determined whether DN XNF-AT could stabilize cytosolic β -catenin protein¹⁸. Notably, high levels of the β -catenin protein were detected when DN XNF-AT was co-expressed with β -catenin (Fig. 4a), thereby indicating that DN XNF-AT stabilizes β -catenin. We then searched for the molecule in the canonical Wnt pathway that

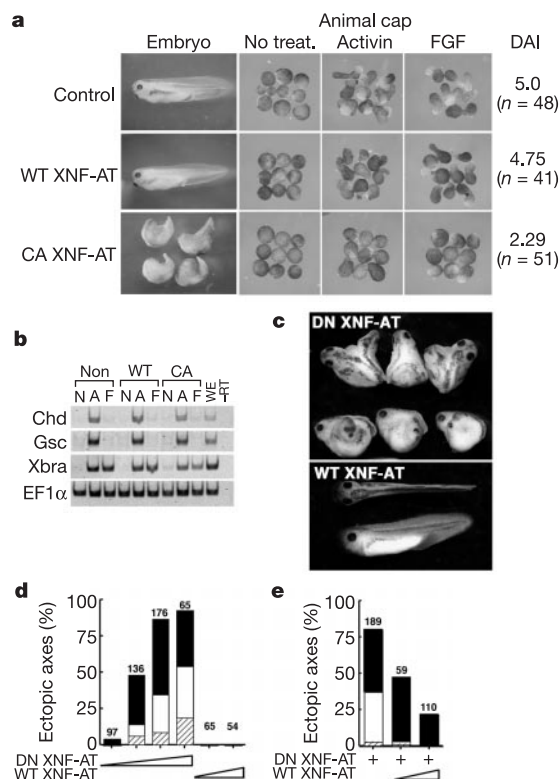


Figure 2 Effects of XNF-AT mutants on dorsoventral axis formation. **a**, Each blastomere at the 4-cell stage was injected with WT XNF-AT (31 pg) or CA XNF-AT RNA (31 pg) (left). Animal caps were dissected at stage 8 from embryos that did not receive an injection (control), embryos injected with WT XNF-AT (125 pg), or CA XNF-AT (31 pg) RNA, and then treated with activin A (10 ng ml⁻¹) or bFGF (100 ng ml⁻¹). Morphology of explants at stage 17 is shown. Average of dorso-anterior index (DAI)¹⁷ is shown at the right. Normal embryos scored DAI 5. Embryos lacking dorso-anterior structures scored DAI 0. **b**, CA XNF-AT did not inhibit mesoderm induction by activin A or by bFGF. RNA from explants at stage 10.5 was analysed using RT-PCR. Chd, *chordin*; Gsc, *goosecoid*; Non, non-injected; WT, injected WT XNF-AT RNA; CA, injected CA XNF-AT RNA. Lane N, untreated; lane A, activin-A treated; lane F, bFGF treated; lane WE, sibling whole embryo; lane -RT, minus reverse transcription of whole embryo. **c**, DN XNF-AT induced dorsal axis duplication. Dorsal (top) or side (bottom) view of embryos injected with DN XNF-AT RNA (0.5 ng) into both ventral blastomeres at the 4-cell stage, and cultured until stage 33. **d**, Histogram of the frequency of axis duplication induced by injection of increasing doses of DN XNF-AT RNA (0.0625–1.0 ng) or WT XNF-AT RNA (0.25–1.0 ng). **e**, Axis duplication induced by DN XNF-AT RNA was rescued by co-injection of WT XNF-AT in a dose-dependent manner. The ectopic axis containing the cement gland is indicated by a white bar, and the induced partial axis is indicated by a filled bar. Hyper-dorsalized embryos are indicated by a hatched bar. Numbers of embryos examined are shown above bars.

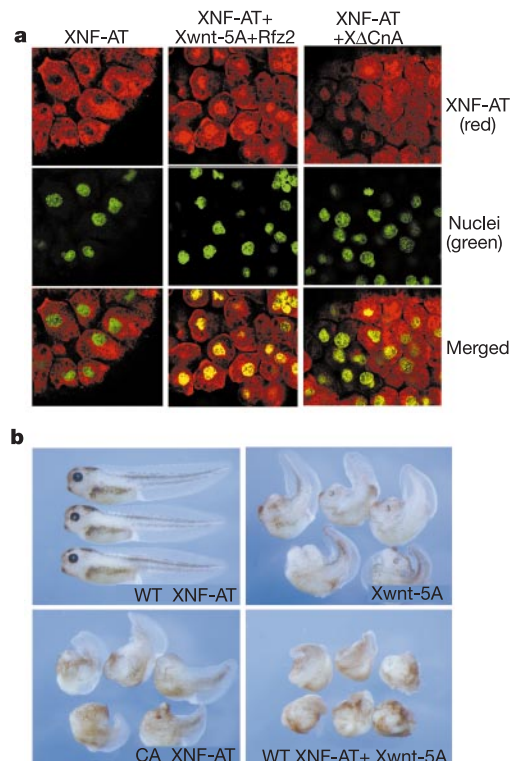


Figure 3 The Wnt/ Ca^{2+} pathway activates XNF-AT signalling. **a**, Subcellular localization of XNF-AT was examined using immunofluorescence analysis of dissected animal cap explants. XNF-AT protein (red) and nuclei (green) were visualized. RNA coding for Myc-tagged WT XNF-AT (0.25 ng), Xwnt-5A (0.25 ng), Rfz2 (0.25 ng) or X ΔCnA (0.25 ng) was injected into both blastomeres at the 2-cell stage. Animal caps were explanted at stages 8–9, and cultured until stage 10. Myc-tagged XNF-AT protein were detected using an anti-Myc antibody, nuclei were stained with Cytochrome (Molecular Probes). **b**, WT XNF-AT enhanced the Xwnt-5A phenotype. RNA coding for WT XNF-AT (0.5 ng), Xwnt-5A (0.1 ng) or CA XNF-AT (31 pg) was injected into both dorsal blastomeres at the 4-cell stage. The embryos were photographed at stage 33.

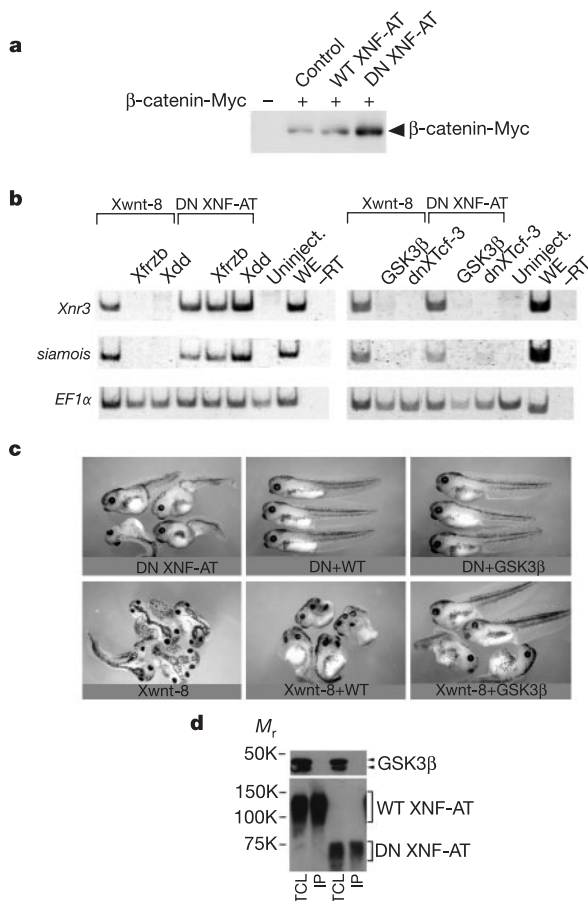


Figure 4 DN XNF-AT activated the canonical Wnt pathway. **a**, DN XNF-AT stabilized β -catenin protein. Blastomeres at the 4-cell stage were injected with Myc-tagged β -catenin RNA (25 pg) alone or with WT XNF-AT RNA (0.5 ng) or DN XNF-AT RNA (0.5 ng), and cultured until stage 10. β -catenin protein was analysed using western blots and an anti-Myc antibody. **b**, DN XNF-AT induced the canonical Wnt-specific markers in the animal cap. Two nanograms of RNA was used, except for Xwnt-8 (10 pg). Lane -RT, minus reverse transcription of the whole embryo. **c**, GSK3 β or WT XNF-AT blocked DN XNF-AT-induced axis duplication. Embryos were injected with DN XNF-AT RNA (0.5 ng) alone or together with GSK3 β RNA (0.5 ng) or WT XNF-AT RNA (0.5 ng) into one ventral blastomere at the 4-cell stage, and cultured until stage 38. **d**, GSK3 β RNA (1 ng) was injected together with Myc-tagged DN XNF-AT RNA (1 ng) or Myc-tagged WT XNF-AT RNA (1 ng), and embryos were collected at stage 7. Lysates were immunoprecipitated using an anti-Myc antibody. The blotted filters were probed with an anti-GSK3 β monoclonal antibody (top; Santa Cruz) or an anti-Myc antibody (bottom). Lane TCL, total cell lysate; IP, immunoprecipitated fraction.

DN XNF-AT might interact with. Figure 4b shows that both the Xwnt-8 antagonist Frzb¹⁹ and dominant negative Xdsh (Xdd²⁰) failed to block expression of the Wnt-specific target genes *siamois* and *Xnr3* that are induced by DN XNF-AT (refs 21, 22). In contrast, *Xenopus* glycogen synthase kinase 3 β (GSK3 β)²³ and dominant negative *Xenopus* T-cell factor 3 (XTcf-3; ref. 24) blocked DN XNF-AT- as well as Xwnt-8-induced *siamois* and *Xnr3* expression. Therefore, DN XNF-AT apparently functions downstream of Xdsh, but upstream of GSK3 β or XTcf-3. The frequency of the DN XNF-AT-induced secondary axes (72.8%, $n = 114$) was reduced by co-injection with GSK3 β RNA (40.4%, $n = 94$) or WT XNF-AT RNA (34.4%, $n = 61$) (Fig. 4c). Immunoprecipitation experiments show that WT XNF-AT or mutants were not associated with GSK3 β (Fig. 4d) nor with other canonical Wnt components, such as Xdsh, axin, XTcf-3 or β -catenin (T.S., unpublished data). Moreover, over-expression of GSK3 β did not affect the principal subcellular

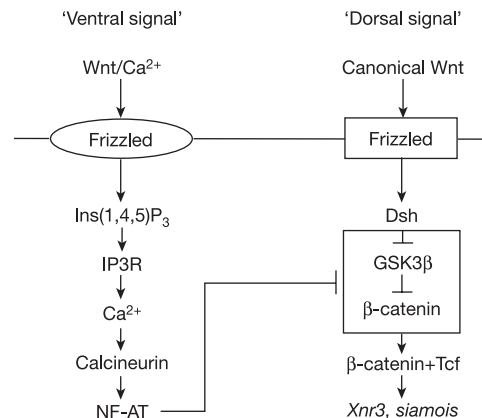


Figure 5 Proposed model for dorsoventral axis formation and the interaction between the Wnt/Ca²⁺ and the canonical Wnt pathways in a *Xenopus* embryo. For the ventralizing signal, the Wnt/Ca²⁺ pathway activates frizzled, then induces a rise in intracellular Ca²⁺ through the Ins(1,4,5)P₃ receptor (IP3R). Elevated intracellular Ca²⁺ activates calcineurin. Activation of calcineurin dephosphorylates NF-AT. Then, NF-AT translocates from the cytosol to the nucleus, and activates expression of a target gene. For the dorsalizing signal, the canonical Wnt pathway activates by means of the maternal Wnt ligand or directly through dorsal accumulation of dishevelled (Dsh). Dsh inhibits GSK3 β -mediated phosphorylation of β -catenin, leading to accumulation of cytosolic β -catenin. β -catenin translocates into the nucleus, binds to T-cell factor (Tcf), and exhibits transcription of *Xnr3* and *siamois*. The target molecule(s) of NF-AT probably have ventralizing activity to inhibit the canonical Wnt pathway downstream of Dsh and upstream of β -catenin. Arrows indicate activation. The T-shaped bar indicates inhibition.

localization of DN XNF-AT (T.S., unpublished data). Given the evidence and that both dominant negative XTcf-3 and GSK3 β blocked DN XNF-AT-induced *Xnr3* or *siamois* expression (Fig. 4b), it is unlikely that GSK3-dependent export of NF-AT²⁵ is the mechanism by which GSK3 β interferes with XNF-AT, although we could not completely rule out the possibility that there was a subtle change in DN XNF-AT subcellular localization by GSK3 β over-expression. Thus, we conclude that the XNF-AT signal interacts with the canonical Wnt pathway downstream of Xdsh and upstream of β -catenin.

As injected *myo*-inositol blocks the effect of dominant negative GSK3 β -induced secondary axis formation²⁶, our findings support the idea that there is cross-talk between phosphatidylinositol cycle signalling and the canonical Wnt pathway. The tyrosine kinase-linked receptor signalling pathway also activates Ins(1,4,5)P₃-Ca²⁺ signalling through phospholipase C γ activation²⁷. FGF receptor tyrosine kinase was shown to trigger calcium influx in *Xenopus*²⁸. Dominant negative FGF receptors had no apparent effect on XNF-AT phosphorylation (T.S., unpublished data); hence, it is unlikely that FGF signalling directly activates XNF-AT signalling. However, XNF-AT might receive inputs from other tyrosine kinase signalling pathways, therefore the activity of XNF-AT in the wild-type embryo may reflect the activity of additional signalling pathways. The proposed model for dorsoventral axis formation and the interaction between the Wnt/Ca²⁺ and the canonical Wnt pathways is shown in Fig. 5. Our evidence suggests that XNF-AT is a direct target of the Ins(1,4,5)P₃-Ca²⁺ signal downstream of the Wnt/Ca²⁺ pathway, and that XNF-AT mediates ventralizing signal by suppression of canonical Wnt activity during axis formation of the *Xenopus* embryo. □

Methods

Isolation of XNF-AT and antibodies

Complementary DNA coding for XNF-AT was isolated by screening a *Xenopus* oocyte cDNA library (gift of D. Melton) using as a probe mouse NF-ATc1 cDNA (from A. Rao).

The XNF-AT antiserum was raised against a peptide (CGWEDLDFRAIFGEE; amino acids 6–20 of XNF-AT). Antiserum was affinity purified on Affigel 15 (Bio-Rad) coupled

with antigen peptide, following the manufacturer's instruction. We used the anti-Myc antibody 9E10.

Embryo manipulation

Embryo manipulations were performed as described². We carried out staging according to ref. 29.

Luciferase assay

For *Xenopus* assays, pNFAT72luc¹¹ and pRL-CMV (Promega), 25 pg per embryo, were injected together with various *in vitro*-transcribed RNAs into four blastomeres of 4-cell stage embryos. After injection, five to eight embryos per group were collected at stage 11 and homogenized to prepare an embryo extract with 1× passive lysis buffer (20 µl per embryo; Promega). Luciferase activity was measured according to protocols for the Dual luciferase assay system (Promega), and analysed using the Model TD-20/20 luminometer (Turner Designs). The data are presented as relative luciferase activity (RLU) normalized by amounts of protein and Renilla luciferase activity in lysates (mean ± s.e.m., *n* = 3). Protein concentration was determined using the Bradford dye assay (Bio-Rad) with bovine serum albumin as the standard.

Analysis of RNA by RT-PCR

Preparation of total RNA and polymerase chain reaction with reverse transcription (RT-PCR) assays was done as described². We used *EF1α* to monitor RNA recovery. Primer sequences were obtained from the *Xenopus* Molecular Marker Resource (<http://cbrmed.ucalgary.ca/pvize/html/WWW/Welcome.html>).

Received 7 November 2001; accepted 21 February 2002.

- Kume, S. Role of the inositol 1,4,5-trisphosphate receptor in early embryonic development. *Cell. Mol. Life Sci.* **56**, 296–304 (1999).
- Kume, S. *et al.* Role of inositol 1,4,5-trisphosphate receptor in ventral signalling in *Xenopus* embryos. *Science* **278**, 1940–1943 (1997).
- Ault, K. T., Durmowicz, G., Galione, A., Harger, P. L. & Busa, W. B. Modulation of *Xenopus* embryo mesoderm-specific gene expression and dorsoanterior patterning by receptors that activate the phosphatidylinositol cycle signal transduction pathway. *Development* **122**, 2033–2041 (1996).
- Kühl, M., Sheldahl, L. C., Park, M., Miller, J. R. & Moon, R. T. The Wnt/Ca²⁺ pathway. *Trends Genet.* **16**, 279–283 (2000).
- Torres, M. A. *et al.* Activities of the Wnt-1 class of secreted signalling factors are antagonized by the Wnt-5A class and by a dominant negative Cadherin in early *Xenopus* development. *J. Cell Biol.* **133**, 1123–1137 (1996).
- Rao, A., Luo, C. & Hogan, P. G. Transcription factors of the NFAT family: regulation and function. *Annu. Rev. Immunol.* **15**, 707–747 (1997).
- Gimlich, R. L. & Gerhart, J. C. Early cellular interactions promote embryonic axis formation in *Xenopus laevis*. *Dev. Biol.* **104**, 117–130 (1984).
- Woodland, H. R. & Jones, E. A. The development of an assay to detect mRNAs that affect early development. *Development* **101**, 925–930 (1987).
- Chow, C. W., Rincon, M. & Davis, R. J. Requirement for transcription factor NFAT in interleukin-2 expression. *Mol. Cell. Biol.* **19**, 2300–2307 (1999).
- Northrop, J. P. *et al.* NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
- Tsuruta, L. *et al.* Cyclic AMP inhibits expression of the IL-2 gene through the nuclear factor of activated T cells (NF-AT) site, and transfection of NF-AT cDNAs abrogates the sensitivity of EL-4 cells to cyclic AMP. *J. Immunol.* **154**, 5255–5264 (1995).
- Saneyoshi, T., Kume, S., Natsume, T. & Mikoshiba, K. Molecular cloning and expression profile of *Xenopus* calcineurin A subunit. *Biochim. Biophys. Acta* **1499**, 164–170 (2000).
- Slusarski, D. C., Corces, V. G. & Moon, R. T. Interaction of wnt and a frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **390**, 410–413 (1997).
- Moon, R. T. *et al.* Xwnt-5A: a maternal Wnt that affects morphogenetic movements after overexpression in embryos of *Xenopus laevis*. *Development* **119**, 97–111 (1993).
- Kühl, M., Sheldahl, L. C., Malbon, C. C. & Moon, R. T. Ca²⁺/calmodulin dependent protein kinase II is stimulated by Wnt and frizzled homologs and promotes ventral cell fates in *Xenopus*. *J. Biol. Chem.* **275**, 12701–12711 (2000).
- Olson, D. J. & Gibo, D. M. Antisense wnt-5a mimics wnt-1-mediated C57MG mammary epithelial cell transformation. *Exp. Cell Res.* **241**, 134–141 (1998).
- Kao, K. R. & Elinson, R. P. The entire mesodermal mantle behaves as Spemann's organizer in dorsoanterior enhanced *Xenopus laevis* embryos. *Dev. Biol.* **127**, 64–77 (1988).
- Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β-catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
- Wang, S., Krinks, M., Lin, K., Luyten, F. P. & Moos, J. M. Frzb, a secreted protein expressed in the Spemann organizer, binds and inhibits Wnt-8. *Cell* **88**, 757–766 (1997).
- Sokol, S. Y. Analysis of Dishevelled signalling pathways during *Xenopus* development. *Curr. Biol.* **6**, 1456–1467 (1996).
- Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A β-catenin/XTcf-3 complex binds to the *siamois* promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
- McKendry, R., Hsu, S. C., Harland, R. M. & Grosschedl, R. LEF-1/TCF proteins mediate wnt-inducible transcription from the *Xenopus* nodal-related 3 promoter. *Dev. Biol.* **192**, 420–431 (1997).
- He, X., St-Jeannet, J., Woodgett, J., Varmus, H. & Dawid, I. Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos. *Nature* **374**, 617–622 (1995).
- Molenaar, M. *et al.* XTcf-3 transcription factor mediates β-catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Beals, C. R., Sheridan, C. M., Turck, C. W., Gadner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
- Hedgepeth, C. M. *et al.* Activation of the Wnt signalling pathway: a molecular mechanism for lithium action. *Dev. Biol.* **185**, 82–91 (1997).
- Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nature*

Rev. Mol. Cell Biol. **1**, 11–21 (2000).

- Musci, T. J., Amaya, E. & Kirschner, M. W. Regulation of the fibroblast growth factor receptor in early *Xenopus* embryos. *Proc. Natl Acad. Sci. USA* **87**, 8365–8369 (1990).
- Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (Daudin) (North-Holland, Amsterdam, 1967).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank E. Amaya, G. Crabtree, D. Kimelman, D. Melton, R. T. Moon, M. Kühl, A. Rao, D. Turner and N. Ueno for plasmids, and Y. Etoh for recombinant activin A. We also thank M. Ohara for critical comments on the manuscript, T. Natsume and K. Takei for discussion, and Y. Takeyama for technical assistance. T.S. was supported by grants from JSPS Research Fellowships for Young Scientists.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K. (e-mail: skume@kaijuu.medic.kumamoto-u.ac.jp) or K.M. (e-mail: mikosiba@ims.u-tokyo.ac.jp). The nucleotide sequence of XNF-AT is deposited in the DNA Data Bank of Japan, European Molecular Biology Laboratory and GenBank under accession number AB037471.

Hedgehog regulates cell growth and proliferation by inducing Cyclin D and Cyclin E

Molly Duman-Scheel^{*†}, Li Weng^{*‡}, Shijie Xin^{*} & Wei Du^{*‡}

^{*} Ben May Institute for Cancer Research and Center for Molecular Oncology, and [‡] Committee on Cancer Biology, The University of Chicago, 924 E. 57th Street, Chicago, Illinois 60637, USA

Although mutations that activate the Hedgehog (Hh) signalling pathway have been linked to several types of cancer¹, the molecular and cellular basis of Hh's ability to induce tumour formation is not well understood. We identified a mutation in *patched* (*ptc*), an inhibitor of Hh signalling, in a genetic screen for regulators of the Retinoblastoma (Rb) pathway in *Drosophila*. Here we show that Hh signalling promotes transcription of *Cyclin E* and *Cyclin D*, two inhibitors of Rb², and principal regulators of the cell cycle during development in *Drosophila*. Upregulation of *Cyclin E* expression, accomplished through binding of Cubitus interruptus (Ci) to the *Cyclin E* promoter, mediates the ability of Hh to induce DNA replication. Upregulation of *Cyclin D* expression by Hh mediates the distinct ability of Hh to promote cellular growth. The discovery of a direct connection between Hh signalling and principal cell-cycle regulators provides insight into the mechanism by which deregulated Hh signalling promotes tumour formation.

During eye development in *Drosophila*, initiation of neural differentiation, marked by an indentation referred to as the morphogenetic furrow, begins at the posterior end of the disc and passes anteriorly. Cells within the furrow arrest in G1 phase before differentiating. Cells located just posterior to the furrow exit G1 arrest and enter a synchronous S phase referred to as the second mitotic wave³. Overexpression of the *Drosophila* Retinoblastoma-family gene (*Rbf*), an inhibitor of the S phase promoting transcription factor E2F^{2,4}, produces a 'rough' adult eye phenotype, characterized by loss of bristles and fusion of ommatidia (Fig. 1b). This phenotype results from delay of S phase progression in cells of the

[†] Present address: Department of Biology, Albion College, Albion, Michigan 49224, USA.

second mitotic wave as a consequence of inhibited E2F target gene expression⁵ (Fig. 1e). Loss of one copy of the *ptc* gene suppresses this rough eye phenotype (Fig. 1c) and restores E2F target gene expression (Fig. 1f). The observed genetic interaction between *Rbf* and *ptc* suggests that the Hh signalling pathway might regulate the cell cycle during eye development.

Hh is secreted from differentiating neurons located just posterior to cells entering S phase in the second mitotic wave (Fig. 1g). This expression pattern is consistent with the idea that reception of the Hh signal might be required for S phase entry in the second mitotic wave. To test this hypothesis, we assessed the effect of blocking Hh signalling during eye development. Cells with a mutated *smoothed* (*smo*) gene cannot respond to the Hh signal⁶ and fail to enter S phase in the second mitotic wave (Fig. 1h). Conversely, when Ci, the transcription factor that mediates Hh signalling⁶, is overexpressed in the furrow, cells normally arrested in G1 enter S phase (Fig. 1i). Ectopic expression of Ci can also promote S phase in G1-arrested cells located in the wing margin (see below) and in the brain⁷. Thus, Ci can induce S phase in a variety of tissues.

Cyclin D, a G1/S cyclin, promotes S phase by inhibiting Rb². During eye development, Cyclin D is expressed in the furrow⁸, and its highest level of expression overlaps with Ci expression (Fig. 2a). This expression pattern is consistent with the idea that Ci protein, which is stabilized in response to reception of the Hh signal from posterior neurons⁶, promotes expression of Cyclin D in the eye. In support of this idea, Cyclin D levels are drastically reduced in *smo* mutant clones that extend through the furrow (Fig. 2b). Conversely, overexpression of Ci induces high levels of *Cyclin D* transcript (not shown) and protein (Fig. 2c) expression in the furrow and in cells immediately surrounding the furrow. In the eye, the ability of Ci to induce high levels of Cyclin D expression is limited to the vicinity of the furrow, the only region where it is capable of inducing S phase. Overexpression of Ci also induces high levels of *Cyclin D* transcript

(Fig. 3a, c) and protein expression (not shown) when expressed ectopically in the wing disc. Thus, in addition to promoting S phase, Ci induces expression of *Cyclin D* in both the eye and wing.

The ability of Hh signalling to induce expression of Cyclin D may explain why increased Hh signalling suppresses phenotypes associated with RBF overexpression (Fig. 1c, f). In support of this idea, overexpression of Ci, which induces Cyclin D expression, also induces ectopic expression of *PCNA*, an E2F target gene, in both the furrow (not shown) and in G1-arrested cells located in the wing margin (Fig. 3g). Coexpression with Ci of RBF-280, a constitutively active form of RBF that cannot be regulated by Cyclin D or Cyclin E⁵, blocks the ability of Ci to induce ectopic *PCNA* expression in the furrow (not shown) and wing margin (Fig. 3i). However, although RBF-280 can block the ability of Ci to induce E2F target gene expression, it does not block the ability of Ci to promote S phase in the eye (Fig. 2g) or wing margin (Fig. 3h, j). These results indicate that although Hh signalling induces E2F target gene expression, it must also be capable of inducing S phase independently of E2F.

A second G1/S cyclin, Cyclin E, is a principal regulator of S phase during *Drosophila* development⁹. Although Cyclin E is an inhibitor of Rb², it also has additional Rb/E2F-independent cell-cycle roles¹⁰. The initiation of high levels of Cyclin E expression in cells of the second mitotic wave, which are located just anterior to neurons secreting Hh protein (Fig. 2d), is consistent with the idea that Hh signalling may regulate Cyclin E expression in the eye. Indeed, loss of *smo* results in reduction of Cyclin E levels in cells entering the second mitotic wave (Fig. 2e). Conversely, when Ci is overexpressed in furrow cells, high levels of *Cyclin E* transcript (not shown) and protein (Fig. 2f) can be detected within these cells. In the eye, the ability of Ci to induce high levels of Cyclin E expression is limited to the furrow region, where it is capable of inducing S phase. Overexpression of Ci is also capable of inducing high levels of *Cyclin E* transcript (Fig. 3b, d) and protein expression (not shown) in the

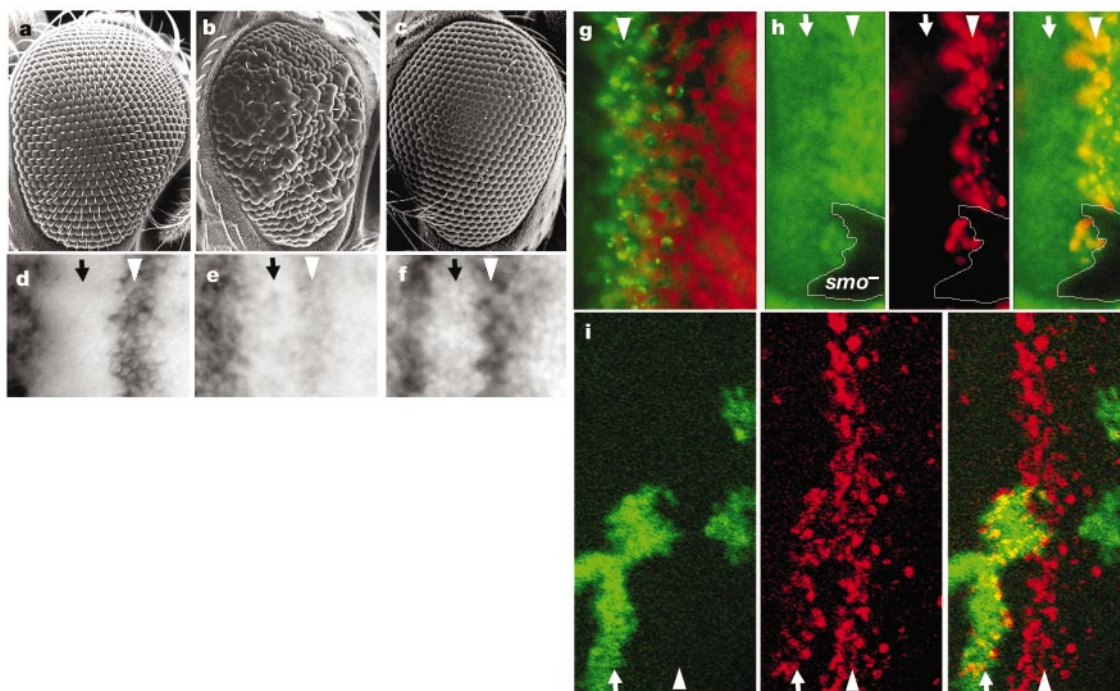


Figure 1 Hh signalling regulates entry into S phase in the developing eye. Here and in subsequent figures, third-instar discs are oriented with the anterior left and dorsal up; arrowheads indicate the second mitotic wave (SMW); arrows indicate the morphogenetic furrow (MF); and overlays are at right, where applicable. **a, b, d, e**, RBF overexpression posterior to the MF induces a rough adult eye phenotype (**b**; wild type, WT, shown in **a**) resulting from loss of E2F target gene expression in the posterior eye during development (**e**, *PCNA* expression; WT in **d**). **c, f**, One copy of the *ptc*^{su53} allele suppresses this rough eye phenotype (**c**) and restores *PCNA* expression (**f**). **g**, DNA replication (green 5-bromodeoxyuridine, BrdU) initiates in the SMW just anterior to cells that express Hh (red; *hh-lacZ* reporter). **h**, Cells within a *smo* mutant clone (loss of GFP expression, left) fail to incorporate BrdU in the SMW (red, centre). **i**, Ectopic BrdU incorporation (red, centre) is detected in a Ci overexpression clone in the MF (green, left).

wing. Although *Cyclin E* is an E2F target gene, the ability of Ci to promote Cyclin E expression in the presence of RBF-280 (Fig. 2h) indicates that Hh can induce Cyclin E expression independently of E2F. Therefore, in addition to promoting expression of Cyclin D and activating E2F, Hh signalling also promotes expression of Cyclin E independently of E2F. The ability of the Cyclin E-dependent kinase inhibitor Dacapo (Dap)^{11,12} to inhibit the ability of Ci to induce S phase in the furrow (Fig. 2i) and wing margin (not shown) indicates that Cyclin E is the principal mediator of the ability of Hh to promote S phase.

To investigate the mechanism by which Hh signalling induces *Cyclin E* transcription, we examined the *Cyclin E* promoter¹³. Several sequences with homology to the consensus Ci-binding site¹⁴ were identified within the 5' regulatory region of *Cyclin E*. Ci was found to bind to three Ci-binding sites (A, B and C; Fig. 4a) in gel shift competition experiments (Fig. 4b). To examine whether this interaction occurs *in vivo*, chromatin immunoprecipitation (ChIP) experiments were carried out. These assays demonstrate that Ci-binding sites A–C are occupied by Ci protein *in vivo* (Fig. 4c). Several lines of evidence indicate that the observed ability of Ci to bind to the *Cyclin E* promoter is important for regulation of *Cyclin E* expression in the developing eye. First, normal upregulation of Cyclin E expression in the second mitotic wave is disrupted in *Cyclin E*^{IP} mutant flies¹⁵ (Fig. 4e), which bear a P element inserted adjacent to Ci-binding sites A–C (Fig. 4a). Also, flies carrying the 16.4 *Cyclin E lacZ*¹³ reporter, which contains all three Ci-binding sites (Fig. 4a), show upregulation of β -galactosidase (β -gal) expression in the

second mitotic wave (Fig. 4e); this pattern resembles the endogenous pattern of Cyclin E expression. By contrast, upregulated β -gal levels in the second mitotic wave are not observed in flies carrying reporter 13.2 *Cyclin E lacZ*¹³ (Fig. 4e), which lacks Ci-binding sites A and B (Fig. 4a). Furthermore, overexpression of Ci in the furrow can drive ectopic β -gal expression from reporter 16.4 (Fig. 4d), but not from reporter 13.2 (not shown). These experiments suggest that the presence of sites A and B is required for normal Hh-mediated upregulation of *Cyclin E* expression in the second mitotic wave. Taken together, these results provide strong evidence that Hh signalling promotes S phase through direct induction of *Cyclin E* expression by the transcription factor Ci.

Cell growth is thought to be regulated independently of cell proliferation¹⁶. For example, overexpression of several well-characterized cell-cycle regulators induces cell proliferation but fails to stimulate growth¹⁷ (defined as the accumulation of mass). In contrast, Cyclin D induces both cell proliferation and promotes the accumulation of mass^{18,19}. Thus, it seems likely that Hh, which induces expression of Cyclin D, may also have a distinct function in regulating cell growth. To test how the modulation of Hh signalling influences growth, we examined the effects of overexpression of Ptc or Ci in clones of undifferentiated wing disc cells. Estimation of clone size was used as a measure of growth, as described previously¹⁷. Ptc overexpression clones are significantly smaller than control clones ($P < 0.0001$; Fig. 5a) and cover only 65% of the area covered by control clones. In contrast, Ci overexpression clones are significantly larger than control clones ($P < 0.0001$; Fig. 5a) and

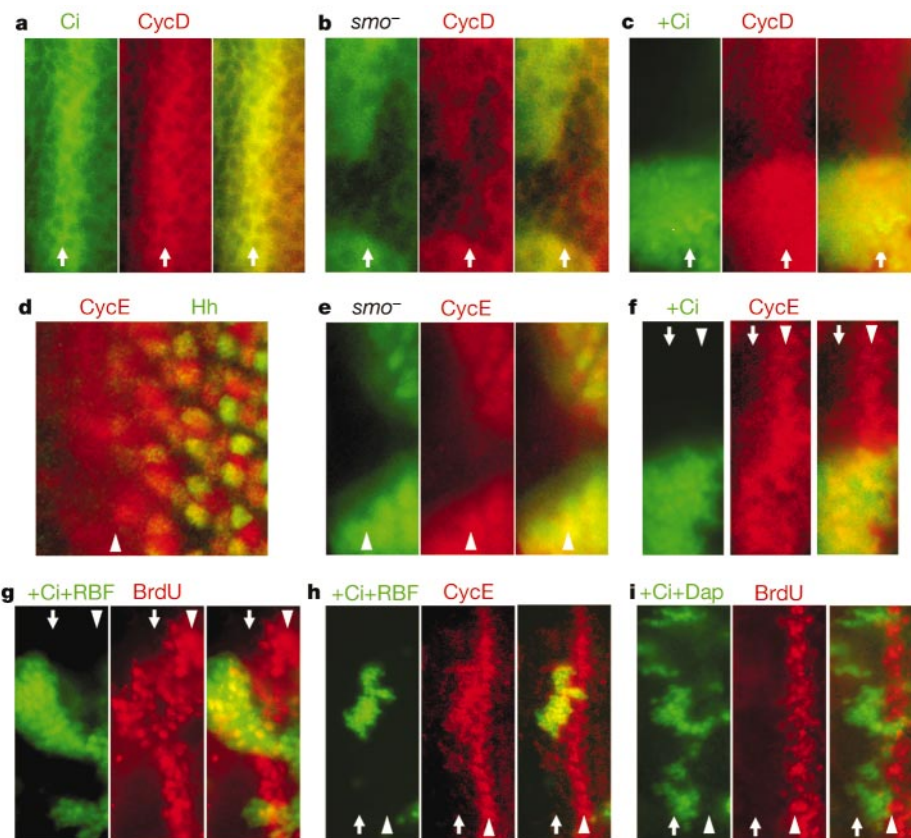


Figure 2 Hh induces *Cyclin D*, *Cyclin E* and E2F-independent S phase in the eye. **a**, In WT discs, expression of Ci (green, left) and Cyclin D (red, centre) overlap in the MF. **d**, High levels of Cyclin E expression in the SMW (red) initiate anterior to *hh-lacZ*-expressing neurons (green). **b**, **e**, *smo* mutant clones (black, left) have decreased levels of Cyclin D (red in **b**, centre) and Cyclin E (red in **e**, centre). **c**, **f**, Overexpression of Ci (green, left)

results in high levels of Cyclin D (red in **c**, centre) and Cyclin E (red in **f**, centre) in the MF. **g**, **h**, Coexpression of RBF-280 with Ci (green clones, left) does not block the ability of Ci to induce S phase (red BrdU in **g**, centre) or Cyclin E expression (red in **h**, centre) in the MF. **i**, Coexpressing Dap with Ci (green clones, left) does block the ability of Ci to induce S phase (red BrdU, centre) in the MF.

cover 143% of the area covered by control clones. These data indicate that, in addition to promoting S phase, Hh signalling has a distinct function in regulating cell growth. The observed ability of Hh signalling to promote cellular growth correlates with previous experiments indicating that Ptc overexpression results in reduced tissue growth²⁰.

It is likely that Cyclin D, which is upregulated upon overexpression of Ci in the wing (Fig. 3a, c), might mediate the ability of Hh to promote growth. Consistent with this hypothesis, overexpression of Ci, like overexpression of Cyclin D–Cyclin-dependent kinase 4 (Cdk4)¹⁸, induces growth (Fig. 5a) and accelerates cell proliferation rates (Fig. 5b) without affecting individual cell size (not shown) or overall cell-cycle phasing (not shown). Likewise, although inhibiting the Hh pathway by overexpressing Ptc decreases growth (Fig. 5a) and decelerates cell proliferation rates (Fig. 5b), it does not affect individual cell size (not shown) or cell cycle phasing (not shown).

We tested the ability of Cyclin D–Cdk4 to mediate induction of growth by Hh in several ways. First, we examined the ability of Cyclin D–Cdk4 to rescue the inhibitory effects of Ptc on growth. The average size of clones expressing Ptc and Cyclin D–Cdk4 is significantly larger than the size of Ptc overexpression clones ($P < 0.0001$; Fig. 5c) and is comparable to the size of clones expressing Cyclin D–Cdk4 alone ($P = 0.07$; Fig. 5c). Thus, overexpression of Cyclin D–Cdk4 can suppress the inhibition of growth by Ptc. Furthermore, although clones of cells lacking Cdk4, the only apparent kinase subunit for *Drosophila* Cyclin D, grow more slowly than wild-type

cells¹⁹, overexpression of Ptc does not significantly reduce the size of clones induced in a Cdk4 mutant background (Fig. 5d). Also, overexpression of Ci cannot stimulate growth in a Cdk4 mutant background (Fig. 5e). These results indicate that Cyclin D–Cdk4 is the principal growth-regulating target of Hh signalling.

The investigation demonstrates that Hh signalling has a distinct ability to promote cellular growth, which is mediated by Cyclin D (Fig. 5). In addition, Hh signalling can induce proliferation during development by promoting expression of *Cyclin D* and *Cyclin E* (Figs 2 and 3). This study reveals a direct connection between Hh signalling and induction of *Cyclin E* expression, which is accomplished through binding of Ci to the *Cyclin E* promoter (Fig. 4). Upregulation of murine *cyclin D1*, *D2* and *E* in response to Hh signalling has been observed²¹. It is therefore likely that the mechanism for *Cyclin E* induction by Hh described here is conserved in mammals. Furthermore, because both overexpression of *Ptc-1* or mutation of *cyclin D1* produces a small mouse phenotype^{22,23}, it is likely that the ability of Hh to promote cellular growth through upregulation of D-type cyclins is also conserved in mice. Thus, constitutive Hh signalling, which promotes deregulated expression of G1/S cyclins that have been associated with diverse forms of human cancer^{24,25}, would promote both cell proliferation and

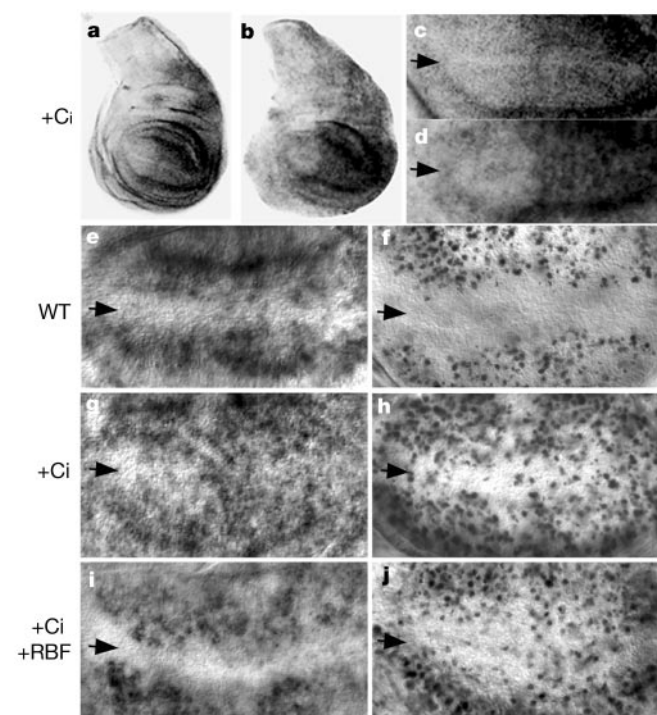


Figure 3 Ci induces *Cyclin D*, *Cyclin E* and E2F-independent S phase in the wing margin. **a–d**, Overexpression of Ci with the *EnGal4* driver induces high levels of *Cyclin D* (**a**, wing margin magnified in **c**) and *Cyclin E* (**b**, wing margin magnified in **d**) mRNA in the posterior of the wing. **e, f**, Normally, *PCNA* mRNA (**e**) and BrdU incorporation (**f**) are not detected in G1-arrested cells³⁰ located in the wing margin. **g, h**, Overexpression of Ci with the wing margin-specific *Gal4C96* driver induces *PCNA* expression (**g**) and BrdU incorporation (**h**). **i, j**, Coexpression of RBF-280 with Ci blocks induction of *PCNA* expression (**i**), but it does not block the induction of S phase by Ci in the wing margin (**j**). Arrows mark the dorsal–ventral boundary. Although the wing was used as a system for testing the effects of Ci overexpression on the cell cycle, these and subsequent experiments are not necessarily meant to suggest that Hh signalling regulates the cell cycle in the wing during normal development.

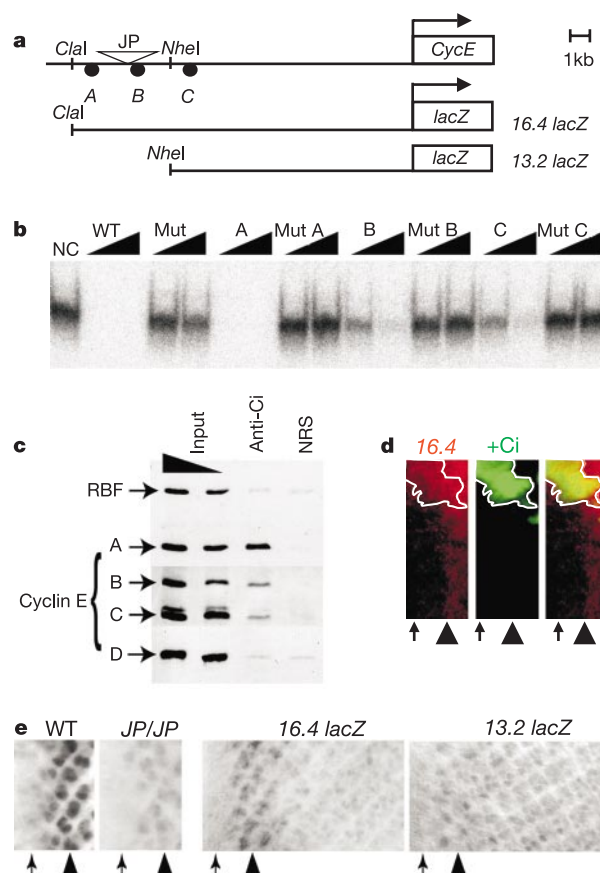


Figure 4 Ci binds to the *Cyclin E* promoter. **a**, The *Cyclin E* promoter region, containing Ci-binding sites A, B and C. Scale is one kilobase. **b**, Unlabelled DNA containing the consensus (WT), A, B or C binding sites competes with labelled consensus oligonucleotides for binding to Ci protein. Mutated (Mut) oligonucleotides served as negative controls. NC, no competition. **c**, ChIP assays indicate that Ci protein occupies binding sites A, B and C *in vivo*. RBF and random fragment D primers are controls. NRS, normal rabbit serum, also a control. **d**, Overexpression of Ci (green, centre) promotes ecotopic β -gal expression from reporter 16.4 in the MF (red, left). **e**, Normal upregulation of Cyclin E expression in the SMW is disrupted in *Cyclin E*^{JP} mutants¹⁵ (JP insertion is indicated in **a**). Upregulation of β -gal expression in the SMW is observed in reporter line 16.4, but not in reporter 13.2. The extent of both reporters is shown in **a**.

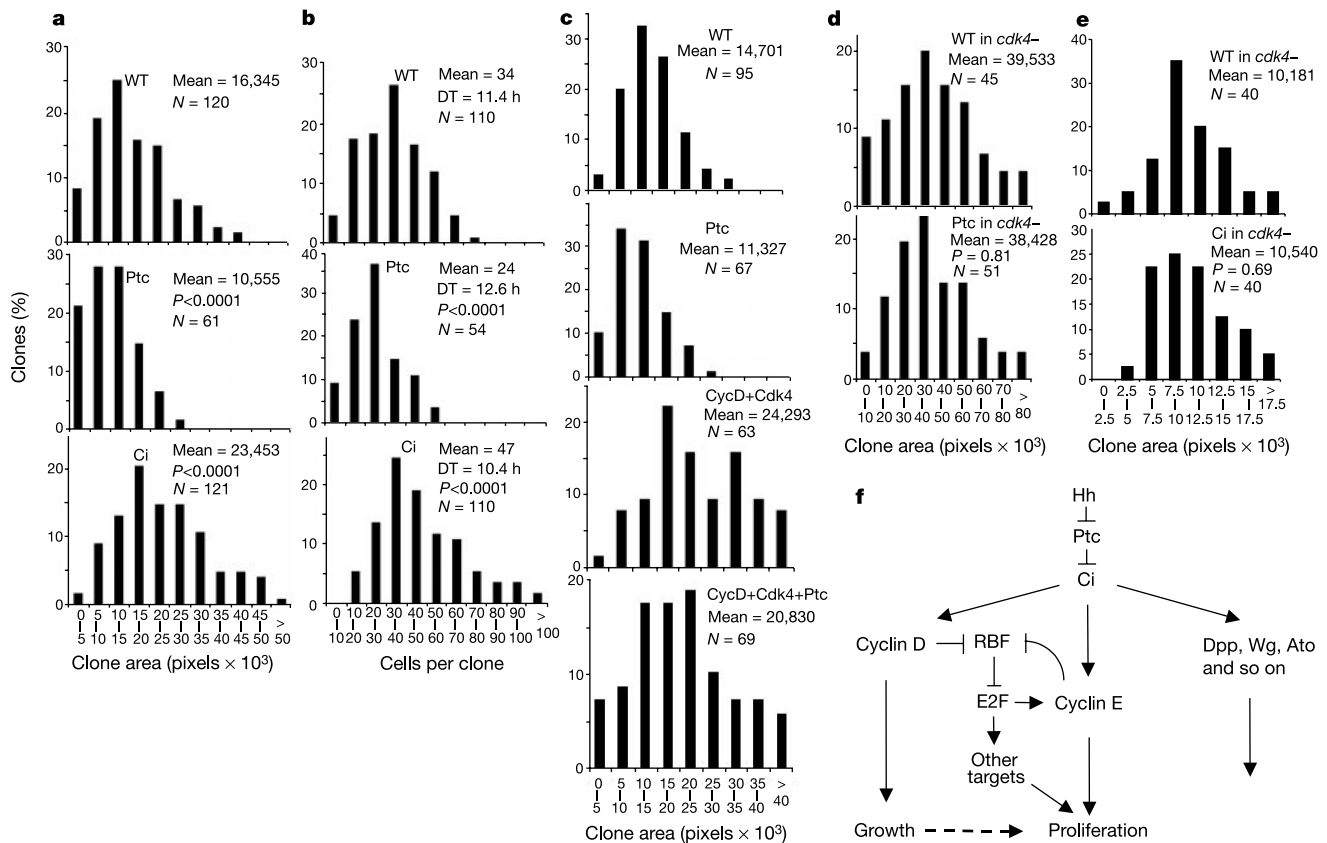


Figure 5 Cyclin D–Cdk4 mediates the ability of Hh to regulate cellular growth. **a**, **b**, Ptc overexpression decreases clone area (**a**) and increases cell doubling time (**b**). Ci overexpression increases clone area (**a**) and decreases cell doubling time (**b**). **c**, Cyclin D–Cdk4 overexpression rescues growth inhibition by Ptc. **d**, **e**, Ptc (**d**) or Ci (**e**)

overexpression do not alter clone size in a *cdk4* mutant background. *P*-values listed indicate comparisons between the experimental group and control in each panel. Other relevant *P*-values are indicated in the text. **f**, A model indicating molecular targets that mediate the effects of Hh on the various cellular processes that occur during development.

growth in tumours. In contrast, during development, cell growth and proliferation must be carefully regulated and coordinated with the processes of cell patterning and differentiation, which are also regulated by Hh signalling⁶ (Fig. 5f). This delicate balance is probably maintained by tight control of the temporal and spatial expression patterns of Hh targets and the molecules that regulate them.

Methods

Further details are available in Supplementary Information.

Fly strains

The following fly strains were used in this study: *GMRRBF^{4C}*, *Cyclin E^{AR95}*, *hh-lacZ*, *P(Ubi-GFP)33,38*, *P(neoFRT40A)*, *smo^{D16}*, *hsFLP²²*, *UAS-ci*, *UAS-GFP^{NLS} S65T*, *Actin 5c > CD2 > Gal4*, *UAS RBF-280*, *UAS-dap*, *En-Gal4*, *Gal4C96*, *Cyclin E^{IP}*, *16.4Cyclin E lacZ*, *13.2Cyclin E lacZ*, *UAS-ptc*, *UAS-Cyclin D*, *UAS-cdk4* and *cdk4³*. More information about these strains is available at Flybase (<http://flybase.bio.indiana.edu>).

Genetic screen

Ethylmethane sulphonate mutagenesis was performed to identify dominant suppressors of the *RBF* overexpression phenotype. *suppressor53 (su53)* mapped to 44C4–44E; complementation tests with known mutations in this region resulted in the identification of *su53* as a *ptc* allele, *ptc^{su53}*.

Generation of clones

The FLP/FRT system²⁶ was used to generate *smo*[−] clones. The flip-out technique²⁷ was used to generate overexpression clones.

Immunohistochemistry and other assays

Immunohistochemistry and *in situ* hybridization were performed essentially as described²⁸. Details of the BrdU incorporation protocol are available in Supplementary Information. Electrophoretic mobility shift¹⁴ and ChIP²⁹ assays were performed essentially as described. Ci-binding sites A–C were mapped with respect to the *CycE^{IP}* insertion by PCR with gene-specific and P1 primers.

Cell growth and proliferation analysis

Cell proliferation and growth rates were performed as described¹⁷. Data were analysed with Student's two-way *t*-test.

Received 31 January; accepted 12 March 2002.

- Ruiz i Altaba, A. Gli proteins and Hedgehog signaling: development and cancer. *Trends Genet.* **15**, 418–425 (1999).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Wolff, T. & Ready, D. F. in *The Development of Drosophila melanogaster* (ed. Bate, M. A. & A. M.) 1277–1326 (Cold Spring Harbor Laboratory Press, Plainview, New York, 1993).
- Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–2262 (1998).
- Xin, S., Weng, L., Xu, J. & Du, W. The role of RbF in developmentally regulated cell proliferation in the eye disc and in Cyclin D/Cdk4 induced cellular growth. *Development* **129**, 1345–1356 (2002).
- McMahon, A. P. More surprises in the Hedgehog signaling pathway. *Cell* **100**, 185–188 (2000).
- Huang, Z. & Kunes, S. Signals transmitted along retinal axons in *Drosophila*: Hedgehog signal reception and the cell circuitry of lamina cartridge assembly. *Development* **125**, 3753–3764 (1998).
- Finley, R. L. Jr, Thomas, B. J., Zipursky, S. L. & Brent, R. Isolation of *Drosophila* cyclin D, a protein expressed in the morphogenetic furrow before entry into S phase. *Proc. Natl Acad. Sci. USA* **93**, 3011–3015 (1996).
- Knoblich, J. A. et al. Cyclin E controls S phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell* **77**, 107–120 (1994).
- Ewen, M. E. Where the cell cycle and histones meet. *Genes Dev.* **14**, 2265–2270 (2000).
- Lane, M. E. et al. Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during *Drosophila* development. *Cell* **87**, 1225–1235 (1996).
- de Noij, J. C., Letendre, M. A. & Hariharan, I. K. A cyclin-dependent kinase inhibitor, Dacapo, is necessary for timely exit from the cell cycle during *Drosophila* embryogenesis. *Cell* **87**, 1237–1247 (1996).
- Jones, L., Richardson, H. & Saint, R. Tissue-specific regulation of cyclin E transcription during *Drosophila melanogaster* embryogenesis. *Development* **127**, 4619–4630 (2000).
- Zarkower, D. & Hodgkin, J. Zinc fingers in sex determination: only one of the two *C. elegans* Tra-1 proteins binds DNA *in vitro*. *Nucleic Acids Res.* **21**, 3691–3698 (1993).
- Secombe, J., Pispis, J., Saint, R. & Richardson, H. Analysis of a *Drosophila* cyclin E hypomorphic mutation suggests a novel role for cyclin E in cell proliferation control during eye imaginal disc development. *Genetics* **149**, 1867–1882 (1998).
- Stocker, H. & Hafen, E. Genetic control of cell size. *Curr. Opin. Genet. Dev.* **10**, 529–535 (2000).
- Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**, 1183–1193 (1998).

18. Datar, S. A., Jacobs, H. W., de la Cruz, A. F., Lehner, C. F. & Edgar, B. A. The *Drosophila* cyclin D–cdk4 complex promotes cellular growth. *EMBO J.* **19**, 4543–4554 (2000).
19. Meyer, C. A. *et al.* *Drosophila* cdk4 is required for normal growth and is dispensable for cell cycle progression. *EMBO J.* **19**, 4533–4542 (2000).
20. Johnson, R. L., Grenier, J. K. & Scott, M. P. *patched* overexpression alters wing disc size and pattern: transcriptional and post-transcriptional effects on *hedgehog* targets. *Development* **121**, 4161–4170 (1995).
21. Kenney, A. M. & Rowitch, D. H. Sonic hedgehog promotes G(1) cyclin expression and sustained cell cycle progression in mammalian neuronal precursors. *Mol. Cell. Biol.* **20**, 9055–90567 (2000).
22. Milenkovic, L., Goodrich, L. V., Higgins, K. M. & Scott, M. P. Mouse *patched1* controls body size determination and limb patterning. *Development* **126**, 4431–4440 (1999).
23. Fantl, V., Stamp, G., Andrews, A., Rosewell, I. & Dickson, C. Mice lacking cyclin D1 are small and show defects in eye and mammary gland development. *Genes Dev.* **9**, 2364–2372 (1995).
24. Donnellan, R. & Chetty, R. Cyclin E in human cancers. *FASEB J.* **13**, 773–780 (1999).
25. Wang, T. C. *et al.* Mammary hyperplasia and carcinoma in MMTV–cyclin D1 transgenic mice. *Nature* **369**, 669–671 (1994).
26. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
27. Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila*. *Cell* **72**, 527–540 (1993).
28. Patel, N. In *In situ Hybridization to Whole Mount Drosophila Embryos* (ed. Krieg, P. A.) 357–370 (Wiley-Liss, New York, 1996).
29. Bosco, G., Du, W. & Orr-Weaver, T. L. DNA replication control through interaction of E2F–RB and the origin recognition complex. *Nature Cell Biol.* **3**, 289–295 (2001).
30. Johnston, L. A. & Edgar, B. A. Wingless and Notch regulate cell-cycle arrest in the developing *Drosophila* wing. *Nature* **394**, 82–84 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Jiang, B. Edgar, C. Lehner, R. Holmgren, H. Richardson, T. Kornberg and the Bloomington Stock Center for providing fly stocks and antibodies used in this study. We gratefully acknowledge C. Heitzig and M. Giorgianni, who performed preliminary experiments that contributed to this work. C. Gudanowski provided technical assistance. We thank E. Williamson and the University of Chicago scanning electron microscope facility, as well as J. Auger and the University of Chicago Flow Cytometry Facility, for their technical advice. We are grateful for discussions with members of the Du lab, the Patel lab, and the Center for Molecular Oncology. Comments from C. Ferguson, A. Mahowald, N. Dyson and J. Duman aided us in the preparation of this manuscript. We thank B. Edgar for communicating results to us before publication. Grants from the American Cancer Society and the National Institutes of Health to W.D. supported this work.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.D. (e-mail: wdu@ben-may.bsd.uchicago.edu).

TREX is a conserved complex coupling transcription with messenger RNA export

Katja Sträßer^{†‡}, Seiji Masuda^{†‡}, Paul Mason^{†§}, Jens Pfannstiel^{*}, Marisa Oppizzi^{*}, Susana Rodríguez-Navarro^{*}, Ana G. Rondón^{||}, Andres Aguilera^{||}, Kevin Struhl[§], Robin Reed[‡], & Ed Hurt^{*}

^{*} BZH, University of Heidelberg, Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany

[‡] Department of Cell Biology; and [§] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

^{||} Department of Genetics, University of Sevilla, E-41012 Sevilla, Spain

[†] These authors contributed equally to this work

The essential yeast proteins Yra1 and Sub2 are messenger RNA export factors that have conserved counterparts in metazoans, designated Aly and UAP56, respectively^{1–7}. These factors couple the machineries that function in splicing and export of mRNA^{1–7}. Here we show that both Yra1 and Sub2 are stoichiometrically associated with the heterotetrameric THO complex⁸, which functions in transcription in yeast^{8–11}. We also show that Sub2

and Yra1 interact genetically with all four components of the THO complex (Tho2, Hpr1, Mft1 and Thp2). Moreover, these components operate in the export of bulk poly(A)⁺ RNA as well as of mRNA derived from intronless genes. Both Aly and UAP56 associate with human counterparts of the THO complex. Together, these data define a conserved complex, designated the TREX ('transcription/export') complex. The TREX complex is specifically recruited to activated genes during transcription and travels the entire length of the gene with RNA polymerase II. Our data indicate that the TREX complex has a conserved role in coupling transcription to mRNA export.

Proteins that interact with Sub2 in yeast were detected by a single-step protein A (ProtA) affinity purification method, which resulted in the identification of Yra1 and several uncharacterized proteins (Fig. 1a, lane 1)⁷. To determine whether the latter proteins specifically associate with Sub2, we employed the more stringent tandem affinity purification (TAP) method for detecting protein interactions, which involves two sequential affinity purification steps¹². Using TAP-tagged Sub2, we detected a significant enrichment of six major proteins that were observed in the original ProtA affinity purification (Fig. 1a, compare lanes 1 and 3)⁷. Consistent with previous studies⁷, one of these proteins is Yra1. The others were a previously unknown open reading frame (ORF) (YNL253w)—hereafter designated Tex1 (for Trex component)—Tho2, Hpr1, Mft1 and Thp2 (Fig. 1a, lane 3). The latter four proteins are all components of the THO complex, which functions in transcription elongation^{8–10,13,14}. Mutations in THO-complex components markedly decrease RNA levels of a *GAL1–lacZ* reporter without affecting promoter function, a phenotype also observed in strains lacking the well-characterized elongation factor TFIIS (ref. 15). In addition, THO-complex mutants are sensitive to the drug 6-azauracil⁹ (K.S., unpublished data), as observed for TFIIS and other transcription elongation factors¹⁵. Finally, THO-complex components interact genetically with Srb2, Srb5, Hrs1, Gal11 and Sin4, which are subunits of the mediator complex required for transcription^{10,13,14,16}.

The observation that a complex that functions in transcription elongation interacts with mRNA export factors raises the possibility that the export factors are recruited to mRNAs by a transcription-coupled mechanism. So we further characterized these interactions. To investigate the specificity of the association between Sub2 and the THO complex, we TAP-tagged Tho2, Hpr1, Mft1 and Thp2, and identified the proteins that are bound in yeast cells. As expected, the other three components of the THO complex co-purified with each of the tagged proteins (Fig. 1a, lanes 4–7). Sub2, Yra1 and Tex1 were also identified in each pulldown assay (Fig. 1a, lanes 4–7). These proteins are approximately stoichiometric with the four components of the THO complex. Moreover, purification of Sub2 from yeast strains lacking single THO-complex components (*hpr1Δ*, *mft1Δ* and *thp2Δ*) causes a complete loss of interaction of the remaining subunits of the THO complex, including Tex1 (see Supplementary Information Fig. 1). We conclude that the THO complex interacts with Sub2 and that Tex1 does not independently interact with Sub2. Finally, the interaction of Sub2, Yra1 and Tex1 with the THO complex is insensitive to RNase (data not shown). We conclude that the THO complex interacts specifically with Sub2 and Yra1 *in vivo*. We have designated this complex TREX because it contains factors involved in both mRNA export and transcription.

To further characterize the TREX complex, we analysed the proteins present after purification from yeast cells by column chromatography, including a stringent 1 M salt elution⁸. All four components of the THO complex were detected on a silver-stained gel (Fig. 1b). In addition, western analysis of this preparation detected Sub2. In contrast, Yra1 was not detected (Fig. 1b). We conclude that Sub2 is more tightly associated with the TREX complex than is Yra1. Thus, Sub2 may mediate the interaction between the THO complex and Yra1.

The mRNA export machinery is conserved between yeast and

humans¹⁷. Thus, we next asked whether the TREX complex exists in humans. Previous pulldown assays from HeLa cell nuclear lysates using glutathione *S*-transferase (GST)–UAP56 identified Aly and additional proteins⁶. Further characterization of these proteins on a low-percentage SDS–polyacrylamide gel revealed two highly specific bands, which migrated at relative molecular mass (M_r) values of about 180,000 (180K) and about 90K (Fig. 1c, lane 4). On a higher-percentage gel, these two bands, a specific band at about 45K, and Aly were detected (Fig. 1d, lane 4). The 180K band

corresponds to a protein whose partial human chromosomal DNA sequence was recently reported as a potential Tho2 homologue¹¹. We constructed a full-length complementary DNA encoding this protein and aligned it with other Tho2 homologues, revealing significant levels of conservation throughout the protein (see Supplementary Information Fig. 2). The 90K band contains peptides that correspond to a new human protein whose full-length cDNA sequence is present in the GenBank database (<http://www.ncbi.nlm.nih.gov/Genbank>). On the basis of the alignment of this protein with other ORFs in the database, the 90K protein is an apparent human homologue of Hpr1 (see Supplementary Information Fig. 3). Finally, the 45K band corresponds to a protein whose full-length cDNA is also present in the database. The closest homologue to this human cDNA encodes the yeast Tex1 protein (see Supplementary Information Fig. 4). Apparent metazoan homologues of Mft1 and Thp2 are not present in the database, and we have not detected them in the UAP56 pulldown assay. We conclude that UAP56 associates with human homologues of Tho2, Hpr1 and Tex1. Consistent with our yeast data, the interaction of UAP56, Aly, human (h) Tho2, hHpr1 and hTex1 was not abolished by RNase treatment of HeLa nuclear lysates (Fig. 1e). We conclude that the TREX complex, containing components of both the mRNA export and transcription machineries, is conserved from yeast to humans. The conservation of this complex suggests that it is critical in mediating interactions between transcription and mRNA export.

Genetic studies in yeast indicate that the biochemical interactions that define the TREX complex are functionally significant. Specifically, we identified all four components of the THO complex in a

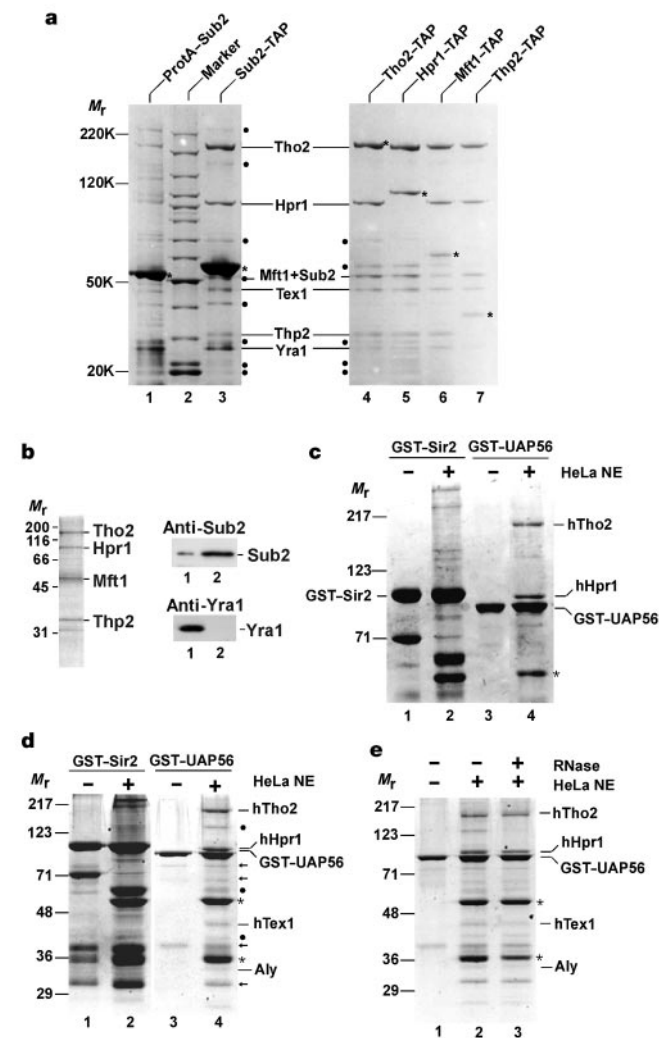


Figure 1 Identification of the conserved TREX complex. **a**, Proteins associated with ProtA–Sub2 (ref. 7) (lane 1) or the indicated TAP-tagged proteins (lanes 3–7) isolated from yeast. Major co-purifying proteins are indicated. Lower-abundance proteins are indicated by filled circles: (from top to bottom) Mlp1 (lane 3), Sac3 (lane 3), Hsp70 (lanes 3–7), Hrb1 (lanes 4–7), Gbp2 (lane 3; co-migrates with Mft1), breakdown product of Sub2 (lane 3) and three ribosomal L-proteins. Asterisks, tagged proteins. **b**, Column chromatography-purified TREX complex analysed by SDS–polyacrylamide gel electrophoresis (PAGE) and silver stained (left) or by western blot using Sub2 and Yra1 antibodies (right). **c**, **d**, control GST–Sir2 (lanes 1 and 2) or GST–UAP56 (lanes 3 and 4) were incubated with buffer (–) or with HeLa nuclear extract (NE) (+). Proteins bound to GST–UAP56 (lane 4) were analysed by 6.5% (**c**) or 10% (**d**) SDS–PAGE. Human (h) Tho2 (AF441770), hHpr1 (Q9UPZ5) and hTex1 (MGC5469) are indicated (see Supplementary Information Figs 2–4). ORF I39463 co-migrates with hHpr1. The lower-abundance proteins are indicated by filled circles: (from top to bottom) Q9BXP4, β -tubulin and MGC2655. Asterisks, abundant lysate proteins. Arrows, breakdown products of GST–UAP56. **e**, GST–UAP56 incubated with buffer (–) or HeLa NE (+), in the absence (–) or presence (+) of RNase.

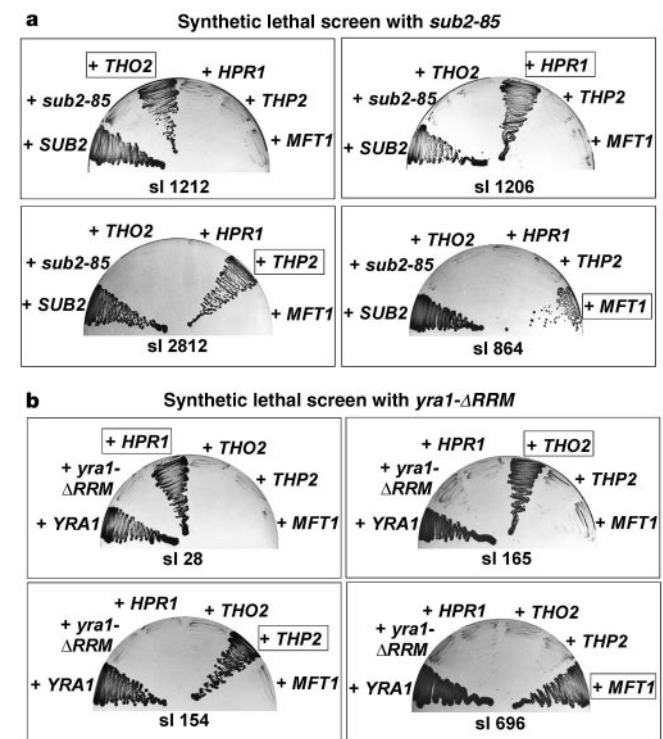


Figure 2 *SUB2* and *YRA1* interact genetically with the genes encoding the four components of the THO complex. **a**, Synthetic lethal screen with the *sub2-85* allele performed at 23 °C. Synthetic lethal (sl) strains 1202, 1206, 2812 and 864 were transformed with plasmids encoding *SUB2*, *sub2-85*, *THO2*, *HPR1*, *MFT1* and *THP2*. **b**, Synthetic lethal screen with the *yra1-ΔRRM* allele performed at 30 °C. Synthetic lethal strains 28, 154, 165 and 696 were transformed with plasmids encoding *YRA1*, *yra1-ΔRRM*, *THO2*, *HPR1*, *MFT1* and *THP2*. Rescue of the synthetic lethal phenotype was assessed by growth after 5 days on plates containing 5-fluoroorotic acid (5-FOA). No growth indicates synthetic lethality.

synthetic lethal screen using a thermosensitive mutant of Sub2, *sub2-85*, which is defective in mRNA export (Fig. 2a)⁷. In this screen, we isolated 28 synthetic lethal mutants. Five of these were complemented by *HPR1*, four by *THO2*, two by *MFT1*, and two by *THP2* (Fig. 2a). This finding is supported by studies showing that *SUB2* is a high copy suppressor of *hpr1* mutants¹⁴.

We next asked whether THO complex components interact genetically with *YRA1*. We previously carried out a synthetic lethal screen with a *yra1* allele, *yra1-ΔRRM*, which lacks the RNA recognition motif (RRM)⁷. This screen led to the identification of Sub2 (ref. 7). Three of the uncharacterized synthetic lethal mutants from this screen were complemented by *THP2* and two by *HPR1* (Fig. 2b). The other two components of the THO complex, *THO2* and *MFT1*, were identified in another synthetic lethal screen with the *yra1-ΔRRM* mutant (Fig. 2b). We conclude that Sub2 and Yra1 interact not only physically but also genetically with the THO complex.

Consistent with our observations linking the THO complex to mRNA export factors, an *hpr1* mutant was previously shown to have an mRNA export defect¹⁸. This defect was detectable because the *hpr1* mutant, like the other THO complex mutants, impairs but does not abolish transcription^{8–10}. To determine whether the other components of the THO complex also have a role in mRNA export, we tested thermosensitive mutants for an mRNA export defect. Some accumulation of poly(A)⁺ RNA was detected in the nucleus at permissive temperature (30 °C), and after a 2-h shift to restrictive temperature (37 °C), a stronger accumulation was observed with the *tho2*, *hpr1*, *mft1* and *thp2* mutants (Fig. 3a). We conclude that all four components of the THO complex function in mRNA export.

Although Sub2/UAP56 was identified as an essential splicing

factor in yeast and metazoans^{19–22}, this protein is required for export of both spliced mRNAs and mRNAs derived from intronless genes^{4–7}. Thus, we next asked whether the THO complex mutants affect export of the heat shock *SSA1* mRNA, which is derived from an intronless gene. To induce transcription of *SSA1* mRNA, wild-type yeast or the THO complex mutants were shifted to the restrictive temperature. *SSA1* mRNA accumulates in the nucleus of each mutant, but not in wild-type cells (Fig. 3b). We conclude that the THO complex is required for efficient export of mRNAs derived from intronless genes.

Studies of the THO components of the TREX complex, together with our observation that mRNA export factors are associated with this complex, raise the possibility that the TREX complex operates in coupling transcription elongation to mRNA export. To address this possibility, we used an *in vivo* assay to determine whether the TREX complex associates with an actively transcribed gene and moves along the gene with RNA polymerase II. We performed a chromatin immunoprecipitation assay in strains containing a long (~8 kilobases) yeast gene (ORF *YLR454*) under the control of the regulatable *GAL1* promoter (Fig. 4a). The strains expressed tagged Hpr1–Myc or Tho2–Myc, from genes integrated into the chromosome. *YLR454* expression was induced by shifting the cells from raffinose- to galactose-containing medium. After induction, the *GAL1* promoter was turned off by addition of glucose to the medium. As shown in Fig. 4b, RNA polymerase II, Hpr1 and Tho2 are not associated with the reporter gene in raffinose (Raf lanes). However, all three proteins associate with the entire region of the transcribing gene in galactose (Gal lanes) as shown with the 5′-, middle-, and 3′-specific polymerase chain reaction (PCR) primers. On addition of glucose, RNA polymerase II, Hpr1 and Tho2 are cleared from the 5′ to the 3′ end of the gene in a time-dependent manner (4-min and 15-min Glu lanes). We conclude that the TREX complex is specifically recruited to the transcribing gene and travels with the polymerase during transcriptional elongation.

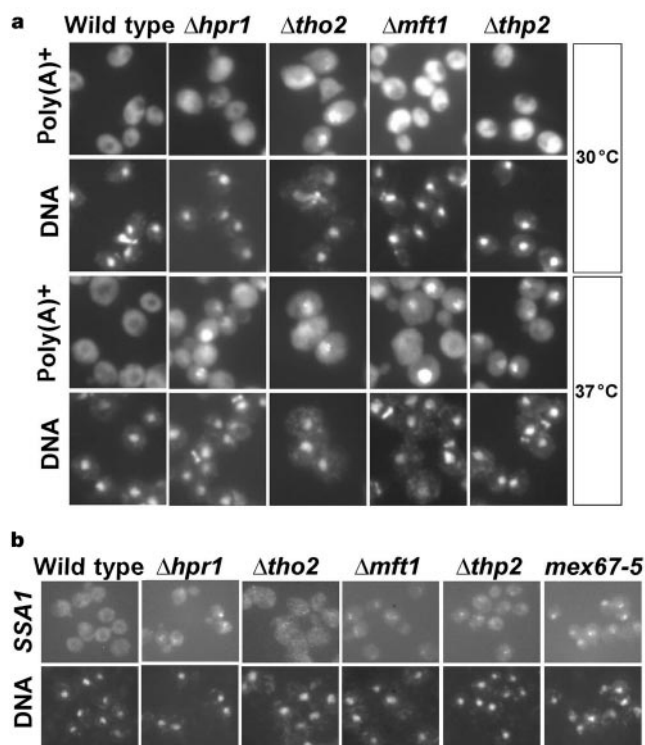


Figure 3 Deletion of *THO2*, *HPR1*, *MFT1* or *THP2* results in an mRNA export defect. **a**, Wild-type yeast (strain W303), $\Delta tho2$, $\Delta hpr1$, $\Delta mft1$ and $\Delta thp2$ strains were grown at 30 °C or shifted for 2 h to 37 °C. Localization of poly(A)⁺ RNA was assessed by *in situ* hybridization. DNA was stained with 4,6-diamidino-2-phenylindole (DAPI). **b**, Accumulation of intronless heat shock *SSA1* mRNA in the nucleus in wild-type, $\Delta tho2$, $\Delta hpr1$, $\Delta mft1$ and $\Delta thp2$ strains, which were grown at 30 °C and shifted for 30 min to 42 °C, as assessed by *in situ* hybridization. The DNA was stained with DAPI.

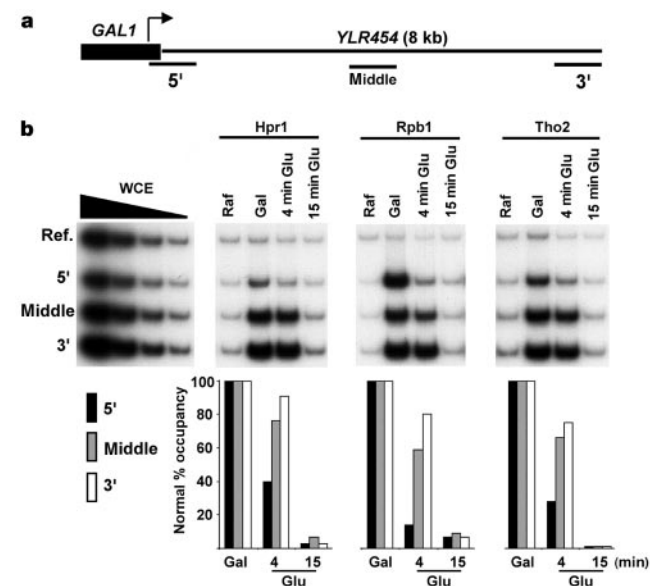


Figure 4 TREX components are recruited to active genes. **a**, Schematic diagram of *GAL1::YLR454* reporter showing primers used for chromatin immunoprecipitation. **b**, *GAL1::YLR454* cells containing Hpr1–Myc or Tho2–Myc were treated with formaldehyde after growth in raffinose followed by glucose addition for 15 min, or after galactose induction followed by 0, 4 or 15 min of glucose incubation. Chromatin was immunoprecipitated with Myc or polymerase II antibodies and PCR was performed using the indicated primers. Immunoprecipitations were quantified and are displayed as the remaining percentage of galactose-induced protein occupancy at each region of the gene at 0, 4 or 15 min after addition of glucose. Ref., reference (see Methods).

The TREX complex physically links proteins that function either in mRNA export or transcription. Two components of the TREX complex, Yra1 and Sub2, are required for mRNA export^{4–7}. A third component, Tex1, is new. Finally, four of the components—Tho2, Hpr1, Mft1 and Thp2—are constituents of the THO complex, which functions in transcription elongation^{8–10,13,14}. The human TREX complex contains Aly, UAP56, hTex1, hTho2 and hHpr1. The remarkable conservation of the TREX complex from yeast to humans reveals the importance of the association of mRNA export and transcription components in the same complex. This conclusion is underscored by our observation that all four components of the THO complex interact genetically with both *YRA1* and *SUB2*. Recent studies in yeast have also identified a link between transcription and mRNA export²³. This work revealed that Npl3 interacts with RNA polymerase II and that both Npl3 and Yra1 are recruited to mRNA during transcription²³. However, the mechanism for this recruitment is not known. Our studies suggest that the conserved TREX complex may be a principal mediator in coupling transcription to mRNA export. In support of this conclusion, our data show that the TREX complex is recruited to actively transcribed genes, where this complex travels with the elongating polymerase along the entire length of the gene. Components of the TREX complex function in export of mRNAs derived from naturally intronless genes as well as bulk poly(A)⁺ mRNA. Thus, it is possible that the TREX complex has a conserved role in loading the mRNA export machinery on to both intron-containing and intron-lacking pre-mRNAs. □

Methods

Yeast analysis

The $\Delta yra1 ade2 ade3$ and the *sub2* shuffle strains were described⁷ as was W303, $\Delta tho2$, $\Delta hpr1$, $\Delta mft1$ and $\Delta thp2$ ¹⁰. The $\Delta sub2 ade2 ade3$ strain for the synthetic lethal screen with *sub2-85* was created by mating the *SUB2* shuffle strain to strain JB²⁴ and tetrad analysis. The TAP tag was integrated into the genome carboxy-terminal of the *SUB2*, *THO2*, *HPR1*, *MFT1* and *THP2* genes by homologous recombination as described¹². Plasmids pUN100-*YRA1* (ref. 1), pRS315-*THO2* (ref. 10), pHT4467-*YRA1*, pRS314-*yra1-ΔRRM* and pRS314-*sub2-85* (ref. 7) were described previously. pHT4467-*SUB2* was constructed by subcloning the *SUB2* ORF containing the *SacI*–*XhoI* fragment of pUN100-*SUB2* into the *SacI* and *XbaI* sites of pHT4467Δ (ref. 24). The synthetic lethal screen and oligo(dT) *in situ* hybridization¹ and *in situ* hybridization with the Cy3-labelled oligonucleotides recognizing the *SSA1* transcript was performed as described⁷.

Protein purification and mass spectrometry

TAP-tagged proteins were purified essentially as described¹². Proteins associated with the purified TAP-tagged proteins were identified by mass spectrometry, which was performed as described²⁴. GST–UAP56 pulldown assays from HeLa nuclear extracts were performed as described⁶. The His-tagged Tho2-purified THO complex was obtained from the strain SchY73 following a described procedure⁶ but using 1 M potassium acetate for elution from the BIO Rex 70 column. Western analysis was performed using polyclonal anti-Sub2 and anti-Yra1 antibodies. Sequence comparison was performed using ClustalW 1.8 (BCM Search Launcher; <http://dot.imgen.bcm.tmc.edu/multi-align/multi-align.html>) and BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html) for printing and shading of multiple alignment files.

Cloning of human Tho2

A partial human cDNA (CAA19741) encoding hTho2 is present in the GenBank database. To obtain a full-length Tho2 cDNA clone, ORF CAA19741 was used to identify XM-047325, which contains the start codon of hTho2. PCR was then used to isolate the full-length human Tho2 cDNA from a HeLa cDNA library using primers 5′-ATGTTTGTTCAGACACAGTTCTAAAGGAA and 5′-GATGGAAGTCTCATATCTGTGCTGTGTC.

Chromatin immunoprecipitation

GAL1::YLR454 cells containing Hpr1-13–Myc or Tho2-13–Myc were grown in synthetic medium containing 2% raffinose. Culture portions (20 ml) were treated for 20 min with 1% formaldehyde, followed by addition of 360 mM glycine. Crosslinked cells were washed twice with cold TBS (20 mM Tris buffer at pH 7.5, 150 mM NaCl). Cell pellets were resuspended in lysis buffer (50 mM HEPES at pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% deoxycholate, 1 mM phenylmethylsulphonyl fluoride) and lysed with a mini bead-beater, followed by sonication sufficient to produce DNA fragments of an average size of 500 base pairs (bp). We used 60% of each soluble extract for immunoprecipitation

with polyclonal anti-Myc antibodies (Upstate Biotechnology), and 15% of each extract was used for immunoprecipitation with anti-Rpb1 antibodies (8WG16, Covance). Immunoprecipitation and PCR were performed as described²⁵. Immunoprecipitated samples and twofold dilutions of whole-cell extract (WCE) were analysed by PCR using primers spanning the *SSA4* promoter ('reference' in Fig. 4b; 300 bp), the *YLR454* translation start codon ('5′' in Fig. 4; 235 bp), the approximate mid-point of the *YLR454* ORF ('middle'; 200 bp), and the *YLR454* translation stop codon ('3′'; 150 bp). The data were quantified by phosphorimager and are displayed as the remaining percentage of galactose-induced protein occupancy at each region of the gene at 0, 4 or 15 min after addition of glucose.

Received 12 March; accepted 16 April 2002.

Published online 28 April 2002, DOI 10.1038/nature746.

1. Sträßer, K. & Hurt, E. C. Yra1p, a conserved nuclear RNA binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* **19**, 410–420 (2000).
2. Stutz, F. *et al.* REF, an evolutionarily conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**, 638–650 (2000).
3. Zhou, Z. L. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401–405 (2000).
4. Gatfield, D. *et al.* The DEXH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr. Biol.* **11**, 1716–1721 (2001).
5. Jensen, T. H., Boulay, J., Rosbash, M. & Libri, D. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr. Biol.* **11**, 1711–1715 (2001).
6. Luo, M.-J. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644–647 (2001).
7. Sträßer, K. & Hurt, E. C. Splicing factor Sub2p is required for nuclear export through its interaction with Yra1p. *Nature* **413**, 648–652 (2001).
8. Chávez, S. *et al.* A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J.* **19**, 5824–5834 (2000).
9. Chavez, S. & Aguilera, A. The yeast *HPR1* gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev.* **11**, 3459–3470 (1997).
10. Piruat, J. I. & Aguilera, A. A novel yeast gene, *THO2*, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J.* **17**, 4859–4872 (1998).
11. West, R. W., Kruger, B., Thomas, S., Ma, J. & Milgrom, E. RLR1 (*THO2*), required for expressing lacZ fusions in yeast, is conserved from yeast to humans and is a suppressor of *SIN4*. *Gene* **243**, 195–205 (2000).
12. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
13. Chang, M. *et al.* A complex containing RNA polymerase II, Paflp, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signalling. *Mol. Cell. Biol.* **19**, 1056–1067 (1999).
14. Fan, H. Y., Merker, R. J. & Klein, H. L. High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Mol. Cell. Biol.* **21**, 5459–5470 (2001).
15. Kulish, D. & Struhl, K. TFIIIS enhances transcriptional elongation through an artificial arrest site in vivo. *Mol. Cell. Biol.* **21**, 4162–4168 (2001).
16. Piruat, J. I., Chavez, S. & Aguilera, A. The yeast *HRS1* gene is involved in positive and negative regulation of transcription and shows genetic characteristics similar to *SIN4* and *GAL11*. *Genetics* **147**, 1585–1594 (1997).
17. Reed, R. & Hurt, E. C. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523–531 (2002).
18. Schneider, R. *et al.* The *Saccharomyces cerevisiae* hyperrecombination mutant *hpr1Δ* is synthetically lethal with two conditional alleles of the acetyl coenzyme A carboxylase gene and causes a defect in nuclear export of polyadenylated RNA. *Mol. Cell. Biol.* **19**, 3415–3422 (1999).
19. Fleckner, J., Zhang, M., Valcarcel, J. & Green, M. R. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP–branchpoint interaction. *Genes Dev.* **11**, 1864–1872 (1997).
20. Kistler, A. L. & Guthrie, C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes Dev.* **15**, 42–49 (2001).
21. Libri, D., Graziani, N., Saguez, C. & Boulay, J. Multiple roles for the yeast *SUB2/yUAP56* gene in splicing. *Genes Dev.* **15**, 36–41 (2001).
22. Zhang, M. & Green, M. R. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev.* **15**, 30–35 (2001).
23. Lei, E. P., Krebber, H. & Silver, P. A. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev.* **15**, 1771–1782 (2001).
24. Baßler, J. *et al.* Identification of a 60S pre-ribosomal particle that is closely linked to nuclear export. *Mol. Cell* **8**, 517–529 (2001).
25. Kuras, L. & Struhl, K. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* **399**, 609–613 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We are grateful to J. Lechner for assistance in the mass spectrometry, J. Bassler for initial help in this project and G. B. Winkler and J. Q. Svejstrup for their help in the isolation of the His-tagged THO complex. E.H. was supported by grants from the Deutsche Forschungsgemeinschaft (SFB, Gottfried Wilhelm Leibniz-Program) and Fonds der Chemischen Industrie; A.A. by grants from the Spanish Ministry of Science and Technology and the Human Frontier Science Program; R.R. and K.S. by National Institutes of Health (NIH) grants. S.M. is supported by the fellowship programme for Japanese scholars and researchers and P.M. is supported by an NIH postdoctoral fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.R. (e-mail: reed@hms.harvard.edu), K.S. (e-mail: Kevin@hms.harvard.edu) or E.H. (e-mail: cg5@ix.uni-heidelberg.de).

The dynamics of actin-based motility depend on surface parameters

Anne Bernheim-Groswasser*, Sebastian Wiesner†, Roy M. Golsteyn*‡, Marie-France Carlier† & Cécile Sykes*

* Laboratoire Physico-chimie 'Curie', UMR 168 CNRS/Institut Curie, 11, rue Pierre et Marie Curie, 75231 Paris cedex 05, France

† Dynamique du Cytosquelette, Laboratoire d'Enzymologie et Biochimie Structurales, UPR A 9063 CNRS, 91198 Gif-sur-Yvette, France

In cells, actin polymerization at the plasma membrane is induced by the recruitment of proteins such as the Arp2/3 complex, and the zyxin/VASP complex^{1–3}. The physical mechanism of force generation by actin polymerization has been described theoretically using various approaches^{4–6}, but lacks support from experimental data. By the use of reconstituted motility medium⁷, we find that the Wiskott–Aldrich syndrome protein^{8,9} (WASP) sub-domain, known as VCA, is sufficient to induce actin polymerization and movement when grafted on microspheres. Changes in the surface density of VCA protein or in the microsphere diameter markedly affect the velocity regime, shifting from a continuous to a jerky movement resembling that of the mutated 'hopping' *Listeria*¹⁰. These results highlight how simple physical parameters such as surface geometry and protein density directly affect spatially controlled actin polymerization, and play a fundamental role in actin-dependent movement.

Investigations of the pathogens *Listeria monocytogenes* and *Shigella flexneri* were important in identifying many proteins that control cellular actin polymerization and cell motility. These bacteria move within a cell by inducing actin polymerization at their surface¹¹ by way of the activation of the Arp2/3 complex^{12,13}. Immunolocalization studies of the Arp2/3 complex in the lamellipodium¹⁴ show that Arp2/3 multiplies the filaments by branching them in a dendritic array¹⁵, as confirmed by biochemical experiments^{16,17}. More recently, surfaces coated with proteins that induce actin polymerization were used to identify new molecular partners for actin polymerization and assembly¹⁸, or to visualize microscopically the actin network formed at the surface¹⁹.

Here we consider how the surface concentration of Arp2/3 complex activator, or the curvature of the surface on which filament branching and growth occur, can affect and control actin-based movement. We have used beads of various diameters (1–10 µm) grafted with the VCA fragment (VCA indicates verprolin-homology, cofilin-homology, and acidic regions), which constitutively activates Arp2/3 complex. We placed the beads in a medium containing a minimal number of purified proteins that support the movement of *L. monocytogenes* or *S. flexneri*⁷. After a certain time that depends on their size, the beads generated behind them a comet (tail) made of actin filaments and started moving. We identified three main characteristic regimes of motion: continuous (smooth), intermittent or erratic, and periodic or saltatory movement. We believe that this is the first evidence that transition from

continuous to saltatory movement can be achieved simply by a change in a physical parameter—that is, the surface density or the bead diameter.

The VCA-coated beads¹⁸ (1–10 µm in diameter) all had the same saturating surface density C_s of VCA, as confirmed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (data not shown). We placed the beads in the mixture of purified proteins⁷, and found that the bead size had a significant effect on the movement type and the comet structure (Fig. 1). Small beads (1–2 µm diameter) moved steadily at a constant velocity (Fig. 1a), as reported in other experimental systems²⁰. Unexpectedly, larger beads displayed a jerky motion. The irregularities in the movement are illustrated by the irregular density (grey level) of the comet (Fig. 1b and c). Furthermore, beads larger than 4.5 µm diameter moved in a periodic hopping manner (Fig. 1c), very similar to the movement of mutated 'hopping' *Listeria*¹⁰. For intermediate-sized beads (3 µm in diameter), a transient or even chaotic movement was observed: the beads hopped randomly, fluctuating from smooth to jerky movement (Fig. 1b). All beads of the same nominal diameter developed comets and showed the same type of motion when sampling the microscope field of view (typically up to 50 beads), showing the robustness of the system.

In these experiments, the surface density of VCA proteins on beads of different diameter was $C_s = (6 \pm 1) \times 10^{-2}$ molecules nm^{-2} . This value corresponds to a maximal surface concentration of VCA. The capacity of immobilized VCA to activate Arp2/3 was evaluated and compared with the activity of VCA in the soluble form using the fluorescence of pyrenyl-actin to monitor the polymerization process. The assay was carried out with identical amounts of immobilized or soluble VCA in parallel. This proved that the specific activity of VCA was the same in both cases, and that the surface concentration of active VCA was independent of the size of the bead.

Mean velocity values, V_{mean} , of beads moving in the solution of purified proteins are given in Table 1. This mean velocity is linearly dependent on the inverse bead diameter. Beads of 1 µm diameter placed in the mixture of purified proteins or in HeLa cell extracts (prepared as previously described²¹) moved with the same average velocity. But in purified protein solution, 100% of the beads started moving by breaking the surrounding symmetrical actin gel ('symmetry breaking') within a few minutes, whereas in the case of cell extracts, symmetry breaking occurred only after 20 minutes and for less than 10% of the beads. Moreover, the velocity fluctuations in the movement of a single bead were 70% in purified proteins as compared to 15% in cell extracts. This can be related to the lower viscosity of the purified protein mixture, as shown in a recent

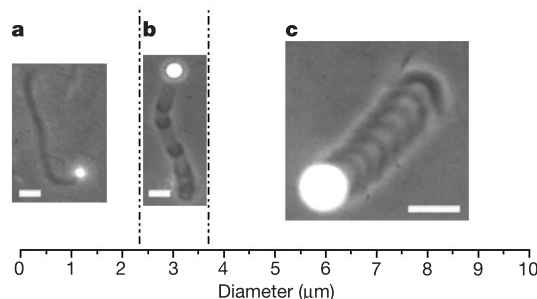


Figure 1 The three main regimes of motion of the beads as a function of bead diameter. In all cases, the beads have a saturated surface concentration of VCA. The dashed vertical lines mark the approximate boundaries between these regimes. **a**, Small beads (<2.5 µm diameter) with a comet of constant gel density (grey level) exhibit continuous motion. **b**, Medium-size beads of 3 µm diameter move in an intermittent manner, as manifested by the irregular gel density of the comet. **c**, Large beads (≥4.5 µm diameter) progress in periodic fashion, as reflected by the alternating comet grey level. Scale bars: **a** and **b**, 5 µm; **c**, 10 µm.

‡ Present address: Division de Cancérologie Expérimentale, Institut de Recherches Servier, 125, chemin de Ronde, 78290 Croissy-sur-Seine, France.

quantitative analysis of *Listeria* movement in fibroblasts²². Beads of diameter $>3\text{ }\mu\text{m}$ placed in cell extracts only rarely produced a comet, and could not be statistically characterized.

Using video phase-contrast optical microscopy, we measured two quantities as a function of time t : the curvilinear distance $D(t)$ from the end of the comet to the bead centre, and the velocity (the time derivative $dD(t)/dt$). These two quantities are respectively plotted in Fig. 2a and b for a $4.5\text{-}\mu\text{m}$ bead at C_s moving in a periodic fashion. Even though actin depolymerization occurs, the end of the comet remained visible until the end of the experiment. The beads that move periodically display a periodic comet structure (Fig. 1c). Analysis of the grey-level intensity of the actin gel along the comet as a function of the distance $X = D(t) - R$, where R is the bead radius, was carried out on a final image. The time dependence of the grey intensity was deduced (Fig. 2c). A correlation was observed between velocity and grey-scale intensity along the comet (Fig. 2b and c, respectively). The velocity of the bead is maximal when the gel density is minimal (low grey level) and vice versa. This correlation can be explained by means of an elastic analysis^{5,23} (see Supplementary Information). In this analysis, a friction force F_f opposing a propulsion force F_e , results from the connection of the gel (via VCA molecules) to the surface of the bead. The model shows that if the molecular connections can break, the friction will depend non-monotonically on the velocity^{5,23}. Consequently, saltatory and continuous regimes are predicted.

The characteristic distance ΔD_{cycle} determined by fast Fourier transform (FFT) analysis of the grey-scale intensity as a function of X is $5.0 \pm 0.8\text{ }\mu\text{m}$, which corresponds to the average size of a step (Fig. 2a). The characteristic time ΔT_{cycle} (also determined by FFT

Table 1 Diameter and mean velocity of beads

Diameter (μm)	V_{mean} ($\mu\text{m min}^{-1}$)
0.954 ± 0.028	1.1 ± 0.2
2.134 ± 0.039	0.63 ± 0.13
2.837 ± 0.133	0.54 ± 0.11
4.537 ± 0.236	0.49 ± 0.12
6.348 ± 0.455	0.34 ± 0.07
9.14 ± 0.709	0.29 ± 0.04

Mean velocity values, V_{mean} , as a function of bead diameter. The values are measured on about 10 beads as the ratio of the curvilinear accumulated distance over time. For beads with diameter $>3\text{ }\mu\text{m}$, the velocity has been averaged over 1.5 h. A linear dependence was found between the mean velocity and the inverse diameter of the beads.

analysis) is 10.7 ± 2.0 minutes. This time is independent of the bead diameter within experimental error, in contrast to ΔD_{cycle} that decreases with the bead size: for $6\text{ }\mu\text{m}$, $\Delta D_{\text{cycle}} = 3.8 \pm 0.3\text{ }\mu\text{m}$, whereas for $10\text{ }\mu\text{m}$, $\Delta D_{\text{cycle}} = 2.5 \pm 0.4\text{ }\mu\text{m}$. For $6\text{-}\mu\text{m}$ -diameter beads and above, the uncertainty in ΔD_{cycle} is rather small (about 10%) and reflects a better-defined period than for $4.5\text{-}\mu\text{m}$ -diameter beads. Stepping behaviour at a molecular scale (5.4 nm) was reported for *Listeria*²⁴, and these authors suggest that it may reflect a highly coordinated progress along many filaments, or discrete events of individual actin filament attachment/detachment to the bacterial surface, which is more likely. We believe that the macroscopic steps on the micrometre length scale observed in our experiments represent a phenomenon that results from a coordinated collective process of a large number of connected (branched) filaments. The two phenomena may act in concert, but our experimental set-up cannot detect discrete molecular events.

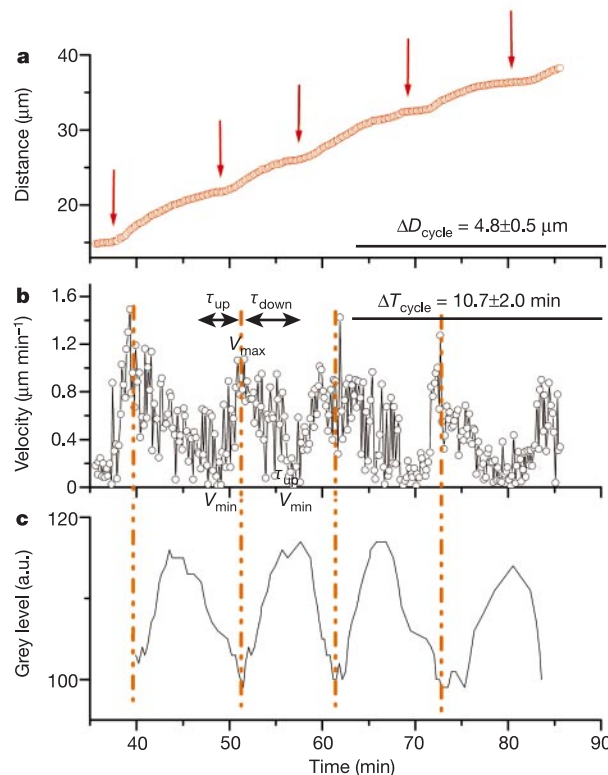


Figure 2 The characteristic behaviour of a $4.5\text{-}\mu\text{m}$ bead at a saturated VCA surface density C_s . The bead exhibits a periodic movement as shown. **a**, The travelled distance (from the comet end to the bead centre) as a function of time shows stepping behaviour. The arrows mark the time of minimum velocity (V_{min}). **b**, The velocity as a function of time, obtained from the time derivative of the travelled distance (**a**), is periodic. During one cycle, the bead accelerates from V_{min} to V_{max} in a time τ_{up} and decelerates to V_{min} with a longer time τ_{down} . The distance accumulated during a step (a jump), ΔD_{cycle} , is

$5.0 \pm 0.8\text{ }\mu\text{m}$, averaged over ten beads, the error representing the standard deviation. The characteristic cycle time ΔT_{cycle} is $10.7 \pm 2.0\text{ min}$, where the error is determined by the full-width at half-maximum of the FFT curve. This time was found to be independent of the bead diameter. **c**, The grey level of the generated actin gel behind the bead as a function of time is shown. The dashed lines show the correlation between minimum grey level (low gel density) and maximum bead velocity.

We studied the dependence of the comet curvature K on the bead diameter D at constant surface density C_s of VCA (see Supplementary Information). We found a power-law dependence, with K proportional to D^{-1} (1.1 ± 0.1). If we consider that the density of actively pushing filaments is constant for all beads, an analysis²⁵ of the bead trajectories based on the Elastic Brownian Ratchet (EBR) model predicts a power of -2 , which is not in agreement with our results. A possible reason for this discrepancy is the assumption²⁵ that each pushing filament is randomly placed on the surface, whereas in practice the filaments are connected and their actions can be correlated.

An early step in actin-based motility of spherical objects in cells, such as lysosomes and endocytic vesicles^{26,27}, is the breakage of the symmetrical actin gel that surrounds the bead²⁸. We characterized symmetry breaking by measuring the time from the instant the beads were placed into the motility medium until they started to move. This time increased linearly with the diameter of the beads (Fig. 3). This linearity can be understood by considering that the actin gel is under tension. Owing to spherical geometry, the growing actin gel is stretched, which results in the development of a lateral stress, $\sigma_{\theta\theta}$, that is maximal at the outer surface of the gel. In turn, the stretched gel exerts a normal stress on the bead²¹. $\sigma_{\theta\theta}$ depends on the Young's modulus C , the thickness of the gel, e , and the radius of the bead, R , as $\sigma_{\theta\theta} = Ce/R$. This relation is still valid at the time the gel fractures. If the thickness of the gel at that time is sufficiently smaller than its thickness in the stationary condition, then e increases linearly with time as a function of the polymerization velocity v_p : $e = v_p t$.

The time of symmetry breaking is given by:

$$t_c = \frac{\sigma_{\theta\theta}^c R}{v_p C} = \alpha R = \frac{\alpha}{2} D$$

where $\sigma_{\theta\theta}^c$ is the critical stress necessary to initiate a fracture; t_c is proportional to the radius, or the bead diameter D . We can estimate the validity of this model by evaluating the slope $a = \alpha/2$. We observed experimentally that the ratio e/R is of the order of 1, and using the relation $\sigma_{\theta\theta}^c = Ce/R$ we deduce that $\alpha = 1/v_p$. The experimental value of the slope a is $2.1 \pm 0.2 \text{ min } \mu\text{m}^{-1}$, thus $\alpha = 4.2 \pm 0.4 \text{ min } \mu\text{m}^{-1}$ and we find that $v_p = 0.24 \pm 0.02 \text{ } \mu\text{m min}^{-1}$. This value is about one order of magnitude smaller than the value we obtained in stress-free pyrene assay measurements, owing to the stress exerted on the surface²¹. The value of t_c changes with the VCA surface density—for example, for $6\text{-}\mu\text{m}$ beads, t_c increases when the VCA concentration decreases. In cells, vesicle size and membrane composition are probably determining factors for symmetry breaking.

To find out how the surface density of VCA controls motility, we tested beads ($4.5\text{-}\mu\text{m}$ in diameter) coated with different concen-

trations of VCA. Decreasing the VCA surface concentration by a factor of 2 (50% saturation of the surface, as determined by SDS-PAGE analysis) induces a transition from saltatory to continuous motion (data not shown). This transition can be related to the drop in the friction force due to the reduction of the surface density of VCA, as qualitatively explained by the elastic analysis^{5,23}. The surface concentration of active VCA was measured using polymerization assay, and was found to be one-sixth of the value under saturating conditions. At concentrations corresponding to a 50-fold decrease of active VCA, no movement was detected. The surface concentration at which the transition occurs was found to depend on the bead diameter.

Although the system described here is far from the complexity of a cell, the use of a limited set of proteins enabled us to identify the physical relationship between surface geometry and the Arp2/3 complex-dependent actin polymerization. It also shows that relevant concentrations of proteins may have important implications for actin-dependent morphogenesis in cell protrusions, as recently shown in cultured cells²⁹. If actin polymerization is capable of deforming a plasma membrane, or, in other words, changing the surface geometry, then the resulting changes in membrane shape can affect subsequent polymerization at the same site—and consequently further out, owing to elasticity. More-complex biochemical changes, such as activation of small GTPases or protein phosphorylation, may act on these inherent physical parameters rather than directly induce the cell movement. Furthermore, we note that changes in physical parameters can mimic the phenotype of the 'hopping' *Listeria*, which was previously induced by altering the ActA protein sequence¹⁰. The ability of physical parameters to reproduce a protein effect greatly enlarges the scope of cellular actin cytoskeleton dynamics, and hence offers more ways to explain the wide range of actin structures within cells at a local level. □

Methods

Derivatization of beads with VCA

VCA was bacterially expressed as a glutathione S-transferase (GST) fusion protein by inserting the corresponding DNA fragments from pSRo-Puro in a modified version of pGEX-4T-1 with an extra *SpeI* site. The stock solution was 1.5 mg ml^{-1} , as determined by Bradford assay.

Polystyrene beads of diameter ranging from 1 to $10\text{-}\mu\text{m}$ were purchased from Polyscience. Beads were incubated in the VCA solution at concentrations in the range of $0.01\text{--}1 \text{ mg ml}^{-1}$ for one hour at room temperature. We used an initial volume of beads that was proportional to the diameter: that is, volumes of $1\text{-}\mu\text{l}$ and $10\text{-}\mu\text{l}$ for beads of $1\text{-}\mu\text{m}$ and $10\text{-}\mu\text{m}$ diameter respectively, given that the volume fraction of the beads in the stock solution was constant, equal to 2.5%. The beads were finally re-suspended in a constant total volume of $20\text{-}\mu\text{l}$ to ensure a constant total surface area of the beads independent of the bead diameter. The beads were stored at 4°C in a storage buffer (20 mM phosphate buffer, pH 7.4, 150 mM NaCl, 1.5 mM NaN_3 , 10 mg ml^{-1} bovine serum albumin (BSA), 5% glycerol) for up to 1 week. We varied the protein concentration on the surface by changing the protein concentration in the incubating solution.

Motility assay

The motility medium contained 8.5 mM HEPES, pH 7.7, 1.7 mM Mg-ATP, 5.5 mM dithiothreitol, 0.12 mM Dabco, 0.1 M KCl, 1 mM MgCl_2 , $6.5 \text{ } \mu\text{M}$ F-actin, $0.1 \text{ } \mu\text{M}$ Arp2/3 complex, $0.046 \text{ } \mu\text{M}$ capping proteins, $2.5 \text{ } \mu\text{M}$ actin depolymerizing factor (ADF), $2.5 \text{ } \mu\text{M}$ profilin, $0.54 \text{ } \mu\text{M}$ α -actinin, 0.31% methyl-cellulose, 0.75% BSA, $1.1 \text{ } \mu\text{M}$ actin-rhodamin as optimized for *Listeria* movement⁷. No vasodilator-simulated phosphoprotein (VASP) was needed. A $0.3\text{-}\mu\text{l}$ volume of bead suspension was added to $20\text{-}\mu\text{l}$ of motility medium. The small volume of the beads ensured that the composition of the motility medium remained unchanged. The sample was placed immediately between a glass slide and a glass coverslip $18 \times 18 \text{ mm}$, sealed with vaseline:lanolin:paraffin (at weight ratio of 1:1:1). To prevent squeezing of the objects, the distance between slide and coverslip was controlled using an inert polyethylene glycol spacer (Goodfellow) to obtain a ratio between sample height and bead diameter of 5 to 6. We found that this ratio ensured that the beads with their comets do not stick to the coverslip walls. In all experiments, symmetry breaking occurs spontaneously, and 100% of the beads generate a comet.

The movement of the microspheres was followed for 1.5 hours by time-lapse optical video-microscopy (phase contrast and fluorescence microscopy, Olympus microscope, $\times 100$). Measurements of bead velocity and the grey-level intensity along the comets were performed using METAMORPH and a software program consisting of a user module (O. Cardoso) for NIH Image (W. Rasband, National Institutes of Health). Both source code and compiled application for MacOS are available at <http://canardo.lbhp.jussieu.fr/cardoso/>. In order to obtain a statistical weight for the data, at the end of the experiment

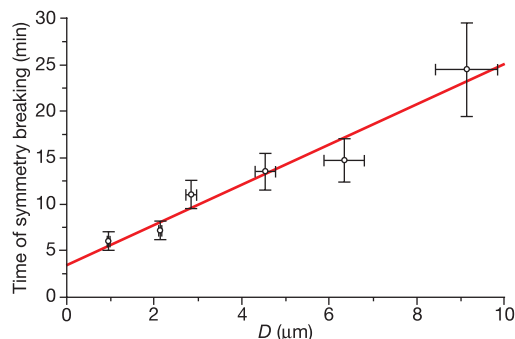


Figure 3 The linear dependence between the time of symmetry breaking and the bead diameter. The slope is $a = \alpha/2 = 2.1 \pm 0.2 \text{ min } \mu\text{m}^{-1}$. The error in the bead diameter is given by the manufacturer, and the error in the time of symmetry breaking is calculated statistically using 10 beads.

we measured the grey-scale intensity of the comets developed by the tracked beads as well as by other ones (about 10).

Determination of activity of immobilized VCA using a polymerization assay

VCA activity was measured as the maximal rate of polymerization of pyrenyl-actin (2.5 μ M) in the presence of Arp2/3 complex (20 nM) and VCA at different concentrations as described¹³. A calibration curve was first established using known concentrations of soluble VCA. The polymerization assays were then performed by adding immobilized VCA in the fluorescence cuvette. 16% sucrose (w/v) was included in the polymerization buffer to prevent sedimentation of the polystyrene beads. The concentration of active immobilized VCA was determined by comparison of the rates measured using VCA-derivatized beads with the calibration curve as a function of soluble VCA. The concentration of actually immobilized VCA was derived from SDS-PAGE. The specific activity of immobilized VCA was derived.

Received 9 January; accepted 20 March 2002.

- Welch, M. D., Mallavarapu, A., Rosenblatt, J. & Mitchison, T. J. Actin dynamics in vivo. *Curr. Opin. Cell Biol.* **9**, 54–61 (1997).
- Renfranz, P. J. & Beckerle, M. C. Doing (F/L)PPPPS: EVH1 domains and their proline-rich partners in cell polarity and migration. *Curr. Opin. Cell Biol.* **14**, 88–103 (2002).
- Pantaloni, D., Le Clainche, C. & Carlier, M.-F. Mechanism of actin-based motility. *Science* **292**, 1502–1506 (2001).
- Mogilner, A. & Oster, G. Cell motility driven by actin polymerisation. *Biophys. J.* **71**, 3030–3045 (1996).
- Gerbal, F., Chaikin, P., Rabin, Y. & Prost, J. An elastic analysis of *Listeria monocytogenes* propulsion. *Biophys. J.* **79**, 2259–2275 (2000).
- Carlsson, A. E. Growth of branched actin networks against obstacles. *Biophys. J.* **81**, 1907–1923 (2001).
- Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M.-F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).
- Machesky, L. M. & Insall, R. H. Scar1 and the related Wiskott-Aldrich syndrome protein, WASP regulates the actin cytoskeleton through the Arp2/3 complex. *Curr. Biol.* **8**, 1347–1356 (1998).
- Takenawa, T. & Miki, H. WASP and WAVE family proteins: key molecules for rapid rearrangement of cortical actin filaments and cell movement. *J. Cell Sci.* **114**, 1801–1809 (2001).
- Lasa, I. *et al.* Identification of two regions in the N-terminal domain of ActA involved in the actin comet tail formation by *Listeria monocytogenes*. *EMBO J.* **16**, 1531–1540 (1997).
- Tilney, L. G. & Portnoy, D. A. Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J. Cell Biol.* **109**, 1597–1608 (1989).
- Welch, M. D., Iwamatsu, A. & Mitchison, T. J. Actin polymerisation is induced by Arp2/3 protein complex at the surface of *Listeria monocytogenes*. *Nature* **385**, 265–268 (1997).
- Egile, C. *et al.* Activation of Cdc42 effector N-WASP by the Shigella IcsA protein promotes actin nucleation by Arp2/3 complex resulting in bacterial actin-based motility. *J. Cell Biol.* **146**, 1319–1332 (1999).
- Svitkina, T. M. & Borisy, G. G. Arp2/3 complex and actin polymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *J. Cell Biol.* **145**, 1009–1026 (1999).
- Mullins, R. D., Heuser, J. A. & Pollard, T. D. The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
- Blanchoin, L. *et al.* Direct observation of dendritic actin filaments networks nucleated by Arp2/3 complex and WASP/Scar proteins. *Nature* **404**, 1007–1011 (2000).
- Pantaloni, D., Boujemaa, R., Didry, D., Gounon, P. & Carlier, M.-F. The Arp2/3 complex branches filament barbed ends: functional antagonism with capping proteins. *Nature Cell Biol.* **2**, 385–391 (2000).
- Fradelizi, J. *et al.* ActA and human zyxin harbour Arp2/3-independent actin-polymerisation activity. *Nature Cell Biol.* **3**, 699–707 (2001).
- Cameron, L. A., Svitekina, T. M., Vignjevic, D., Theriot, J. A. & Borisy, G. G. Dendritic organization of actin comet tails. *Curr. Biol.* **11**, 130–135 (2001).
- Cameron, L. A., Footer, M. J., van Oudenaarden, A. & Theriot, J. A. Motility of ActA protein-coated microspheres driven by actin polymerisation. *Proc. Natl Acad. Sci. USA* **96**, 4908–4913 (1999).
- Noireaux, V. *et al.* Growing an actin gel on spherical surfaces. *Biophys. J.* **78**, 1643–1654 (2000).
- Giardini, P. A. & Theriot, J. A. Effects of intermediate filaments on actin-based motility of *Listeria monocytogenes*. *Biophys. J.* **81**, 3193–3203 (2001).
- Gerbal, F. *et al.* On the 'Listeria' propulsion mechanism. *Pramana J. Phys.* **53**, 155–170 (1999).
- Kuo, S. C. & McGrath, L. Steps and fluctuations of *Listeria monocytogenes* during actin-based motility. *Nature* **407**, 1026–1029 (2000).
- Rutenberg, A. D. & Grant, M. Curved tails in polymerization-based bacterial motility. *Phys. Rev. E* **64**, 21904–21907 (2001).
- Merrifield, C. J. *et al.* Endocytic vesicles move at the tips of actin tails in cultured mast cells. *Nature Cell Biol.* **1**, 72–74 (1999).
- Taunton, J. *et al.* Actin-dependent propulsion of endosomes and lysosomes by recruitment of N-WASP. *J. Cell Biol.* **148**, 519–530 (2000).
- Van Oudenaarden, A. & Theriot, J. A. Cooperative symmetry-breaking by actin polymerisation in a model for cell motility. *Nature Cell Biol.* **1**, 493–499 (1999).
- Beningo, K. A., Dembo, M., Kaverina, I., Small, J. V. & Wang, Y. Nascent focal adhesions are responsible for the generation of strong propulsive forces in migrating fibroblasts. *J. Cell Biol.* **153**, 89–100 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Plastino for discussions and the purification of VCA proteins, and F. Castellano and P. Chavrier for the gift of the plasmid encoding VCA. Theoretical discussions were conducted by J. Prost. We thank D. Didry for the purification of Arp2/3, ADF-cofilin, and actin, and R. Boujemaa for the purification of the capping protein. We thank H. Boukellal

for help in determining the VCA concentration on the beads, E. Paluch for helping in analysing the videos and K. Sekimoto and D. Pantaloni for discussions.

Correspondence and request for material should be addressed to C.S.

(e-mail: cecile.sykes@curie.fr), or M.-F.C. (e-mail: carlier@lebs.caps-gif.fr) for the motility medium.

Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold

Hariharan Jayaram*, Zenobia Taraporewala†, John T. Patton† & B. V. Venkataram Prasad*

* Program in Structural and Computational Biology and Molecular Biophysics, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

† Laboratory of Infectious Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Rotavirus, the major cause of life-threatening infantile gastroenteritis, is a member of the *Reoviridae*¹. Although the structures of rotavirus² and other members of the *Reoviridae*^{3,4} have been extensively studied, little is known about the structures of virus-encoded non-structural proteins that are essential for genome replication and packaging. The non-structural protein NSP2 of rotavirus, which exhibits nucleoside triphosphatase, single-stranded RNA binding⁵, and nucleic-acid helix-destabilizing⁶ activities, is a major component of viral replicase complexes^{7,8}. We present here the X-ray structure of the functional octamer⁹ of NSP2 determined to a resolution of 2.6 Å. The NSP2 monomer has two distinct domains. The amino-terminal domain has a new fold. The carboxy-terminal domain resembles the ubiquitous cellular histidine triad (HIT) group of nucleotidyl hydrolases¹⁰. This structural similarity suggests that the nucleotide-binding site is located inside the cleft between the two domains. Prominent grooves that run diagonally across the doughnut-shaped octamer are probable locations for RNA binding. Several RNA binding sites, resulting from the quaternary organization of NSP2 monomers, may be required for the helix destabilizing activity of NSP2 and its function during genome replication and packaging.

Rotavirus is a large and complex virus. Its genome, enclosed within three concentric capsid layers, consists of 11 segments of double-stranded RNA (dsRNA) that code for six structural and six non-structural proteins (NSPs)¹. Several *in vivo* and *in vitro* studies have strongly implicated NSP2, a highly basic protein of relative molecular mass $M_r = 35,000$ which exists as octamers, in genome replication and packaging. This protein, in association with viral RNA and polymerase, localizes to viroplasm of the infected cells where these processes occur^{7,11,12}. Temperature-sensitive mutants of NSP2 fail to replicate the genome and produce mostly empty particles, which implicates NSP2 in genome replication and packaging^{13,14}. Recent studies on recombinant NSP2 have shown that it is a magnesium-dependent NTPase that binds to ssRNA in a sequence-independent manner⁵ and exhibits nucleic-acid helix-destabilizing activity⁶. On the basis of these studies it is suggested that NSP2 functions as a molecular motor that uses the energy derived from NTP hydrolysis to package the viral genome⁹. In its function, NSP2 may be homologous to NS2 (refs 15, 16) of blue-tongue virus and σ NS of orthoreovirus¹⁷. To understand the structural mechanisms underlying the function of NSP2, we have

determined the X-ray crystallographic structure of His-tagged recombinant NSP2 to 2.6 Å resolution. Given the current status of rotavirus vaccines¹⁸, such structural studies may prove valuable in developing alternative strategies for counteracting this human pathogen, which accounts for ~600,000 deaths annually worldwide.

The NSP2 monomer has two distinct domains, an N-terminal domain from residue 1 to 141 and a C-terminal domain from 151 to 313 (Fig. 1a, b). The C-terminal four amino acids and the 6-His tag are not observed in the crystal structure and are probably disordered. The two domains are separated by a 10-residue loop from residue 141 to 151. A striking feature of the monomer is a 25-Å-deep cleft that lies between the two domains. NSP2 crystallizes as an octamer using crystallographic 4-2-2 symmetry, resulting in one monomer per asymmetric unit.

The N-terminal domain, although predominantly α -helical, has two pairs of anti-parallel β -strands towards the N terminus. This domain can be described as having two sub-domains: the first one, between the residues 1 and 51, consists of two pairs of β -strands separated by two α -helices (α 1 and α 2), and the second, between residues 77 and 141, consists of a cluster of four α -helices (α 3- α 6). These two sub-domains are connected by a 24-residue-long predominantly basic loop (Fig. 1a, arrow). There are three highly conserved prolines (53–55) at the start of this loop, which may be essential for the conformation of this loop (Fig. 1c). This loop lines the prominent grooves that run diagonally across the 2-fold axes of the doughnut-shaped octamer, as described below, and contributes to their basic character.

The C-terminal domain has an α/β fold. The prominent feature

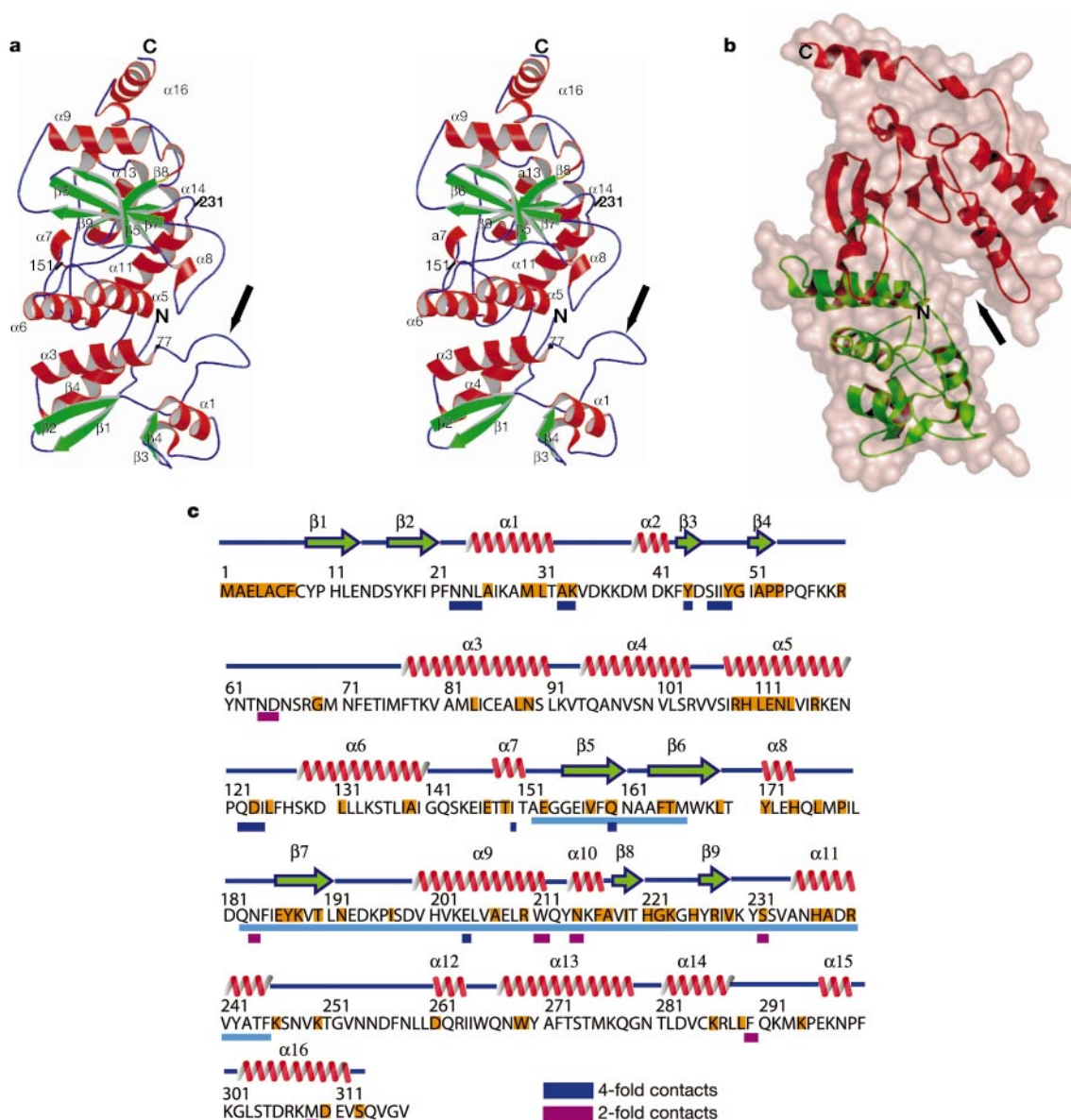


Figure 1 The X-ray structure of the monomeric subunit of NSP2. **a**, A stereo view of the ribbon representation of the structure. The secondary structural elements in the monomer are coloured: α -helices in red, β -strands in green and loops in blue. The locations of the major α -helices are indicated following the same numbering scheme as in **c**. The basic loop between the subdomains in the N-terminal domain is shown by an arrow. **b**, Another view of the monomer showing the cleft (arrow) and the domain organization. The N- and C-terminal domains are shown in green and red respectively. The location of the N and the

C termini are denoted in both **a** and **b**. **c**, The structure–sequence relationship in NSP2. The secondary structural elements along with the sequence are shown following the same colour scheme as in **a**. α -helices and β -strands are numbered sequentially. The strictly conserved residues are highlighted in orange. The residues that participate in the 4- and 2-fold contacts of the octamer are indicated by boxes below the sequence in blue and purple respectively. The HIT-homology (see Fig. 2) region is underlined in light blue.

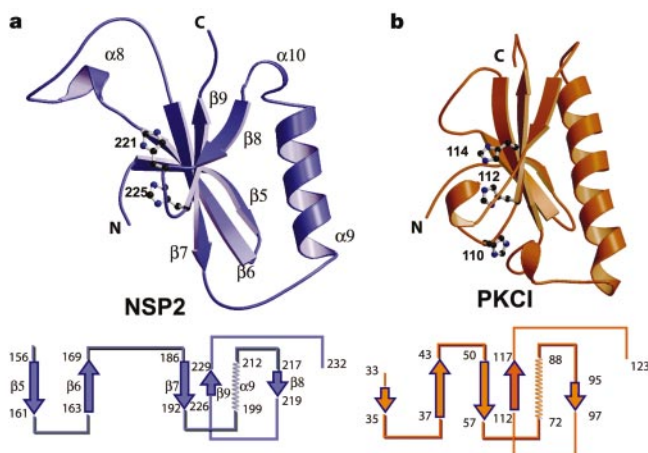


Figure 2 The HIT-like fold in NSP2. **a**, The region of the NSP2 structure which shows the HIT-like fold (above), along with a topology diagram (below). **b**, Corresponding domain in the PKCI (PDB entry code: 1KPF) structure (above), along with the topology diagram (below). The HIT fold typically has $\beta 3\alpha\beta 2\alpha$ configuration. The three His residues of the triad in the PKCI, and two of the histidines in NSP2 are also shown. Structural similarity

was discovered using the DALI server¹⁹, which indicated statistically significant (Z -scores > 3.6) homology with three of the HIT proteins. The r.m.s. deviation in the matching C α atoms ranged from 2.3 to 3.3 Å for these HIT proteins. The PKCI protein, despite having a lower r.m.s. value (3.3 Å for matching 71 C α atoms), had a higher Z -score (4.2).

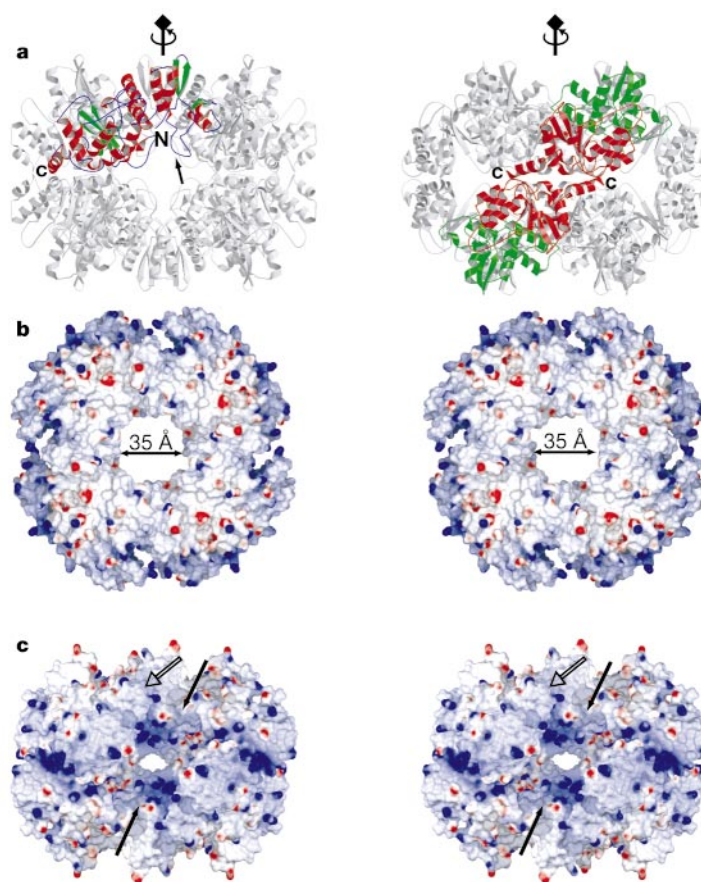


Figure 3 Octameric structure. **a**, Views of the crystallographic octamer along the two distinct 2-fold axes, separated by a 45° rotation about the 4-fold axis. The left side shows a ribbon representation of the octamer along one of these two 2-fold axes. One of the subunits, with N- and C-terminal ends marked, is shown using the same colouring scheme as in Fig. 1a. The rest of the subunits are shown in grey. A deep groove lined by the basic residues from the N-terminal loop, shown by an arrow, runs diagonally across this 2-fold axis. The right side shows the octamer as viewed along the other 2-fold axis of the 4-2-2 symmetry. This 2-fold axis is associated with protein-protein interactions that link the two tetrameric layers. The two subunits that participate in these 2-fold interactions

are highlighted using the colour scheme in Fig. 1b. The N-terminal domains (in green) participate in the 4-fold related inter-subunit interactions, whereas the C-terminal domains (in red) participate in inter-subunit interactions across this 2-fold axis. **b**, Stereo view of the electrostatic potential surface of the octamer viewed along the 4-fold axis showing the central hole. **c**, Stereo view of the octamer, along the 2-fold axis, as in **a** (left) showing the grooves (black arrows). An open arrow shows the location of the cleft in one of the subunits. The surface is coloured deep blue in the most electropositive regions (20 $k_B T$) and red in the most electronegative regions ($-20 k_B T$), where k_B is the Boltzmann constant, and T is temperature in kelvin.

of this domain is the twisted anti-parallel β -sheet flanked by α -helices, formed by residues from 150 to 245, in a $\beta 3\alpha\beta 2\alpha$ configuration. The anti-parallel β -strands, between residues 186 and 191 ($\beta 7$), and between 226 and 230 ($\beta 9$), along with the loop, between residues 221 and 226, constitute the base of the cleft. The inter-domain loop and the helix $\alpha 5$ from the N-terminal domain form a major part of one side of the cleft, whereas the other side of the cleft is made of helix $\alpha 11$ and a loop between residues 245 and 260. The remainder of the C-terminal region, from residues 260 and 311, consists predominantly of α -helices ($\alpha 12$ – $\alpha 16$).

Structural comparison of NSP2 with other protein structures using the DALI server¹⁹ did not show any statistically significant similarity for the N-terminal domain. However, the search indicated a significant homology for the C-terminal domain with the structure of the protein kinase C interacting protein (PKCI), a prototypical member of the HIT family of nucleotide interacting proteins^{10,20} consisting of over 35 proteins from across Archaea, Prokaryae and Eukaryae²¹ (Fig. 2).

Despite this structural similarity, NSP2 has no sequence identity with any of the HIT proteins, and does not have the signature histidine-triad sequence (HxHxHxx). However, when the two structures are optimally superimposed, there is a noticeable match between His 225 of NSP2 and His 112 of PKCI (within 1 Å). The conserved His 112 (second histidine of the triad) in PKCI is the main catalytic residue in the in-line attack on the α -phosphate during hydrolysis¹⁰. In the vicinity of His 225, in NSP2, there are two prominent spheres of ordered density. On the basis of the surrounding residues, and the strength of the density peak (2.8 sigma), one of them is possibly a Mg^{2+} ion coordinated by the side-chain carboxyl- and hydroxyl-oxygen atoms of Glu 153 and Tyr 171 respectively. The base of the cleft in NSP2 also has a cluster of conserved basic residues, which include two histidines (His 221, His 110), an arginine (Arg 227) and a lysine (Lys 188). His 221 matches with His 110 of PKCI (the first histidine of the triad), although not as well as His 225. These features strongly suggest that the active site for NTP hydrolysis is inside the cleft, with one of the histidines perhaps serving as an important catalytic residue. Considering that most HIT proteins hydrolyse at the α -phosphate, it is quite likely that NSP2, which hydrolyses γ -phosphate, functions by a mechanism that does not involve a strict triad of histidines.

In contrast to HIT proteins, which are intimate dimers²¹, NSP2 exhibits an intrinsic ability to oligomerize into octamers, as shown by dynamic light scattering, analytical ultracentrifugation⁹ and preliminary electron cryomicroscopy studies (data not shown). Extensive interactions between the monomeric subunits, with buried surface area of 985 Å² per monomer across the 2-fold axis and 710 Å² per monomer across the 4-fold axis seen in the crystallographic octamer, indicate that NSP2 crystallizes as intact octamers. Comparison of the available 18 sequences of the NSP2 (protein families database (PFAM) code is PF02509; see <http://pfam.wustl.edu>) of various rotavirus strains, in relation to the X-ray structure, shows that several conserved residues participate in the inter-subunit interactions (Fig. 1c). In the formation of the octamer with 4-2-2 symmetry, which has two distinct 2-fold axes perpendicular to the 4-fold axis (Fig. 3a), the N-terminal domain is mainly involved in the interactions around the 4-fold axis, whereas the C-terminal domain is involved in the interactions across one of the two 2-fold axes. A distinguishing feature of the dimeric interactions in the HIT proteins is the extension of the β -sheet across the 2-fold interface. In NSP2, because of the orientations of the 2-fold related subunits in the octamer (Fig. 3a, right) such extended sheet formation is not seen, and thus the similarity between the NSP2 and HIT is limited only to the tertiary structure level.

To understand how NSP2 might bind to RNA, we examined the octameric structure in terms of surface features and electrostatic charge distribution. The NSP2 octamer is a doughnut-shaped particle with a 35 Å central hole (Fig. 3b). On the outer surface,

perpendicular to the 4-fold axis, there are four prominent grooves, each ~30 Å wide and 25 Å deep, that run diagonally across the other 2-fold axes of the 4-2-2 octamer (Fig. 3a left and c). This surface of the octamer is relatively basic, particularly the grooves, which are predominantly lined by basic residues from a loop that links the two sub-domains in the N-terminal domain. The surface surrounding the central hole is not as basic and is relatively hydrophobic (Fig. 3b), perhaps suitable for binding viral polymerase, given that NSP2 interacts with it *in vivo*¹². The dimensions and electrostatic characteristics of the groove make it a likely candidate for the binding of RNA. This proposed binding of nucleic acid on the outside is similar to that observed in the *Escherichia coli* single-strand binding protein (SSB), a tetrameric helix-destabilizing protein, where the ssDNA wraps around the tetramer²². We suggest that the increased stabilization of ssRNA on binding to the basic groove results in NSP2 shifting the dsRNA–ssRNA equilibrium in favour of ssRNA. In SSB, several aromatic residues are involved in high-affinity binding to ssDNA. In NSP2, conserved and exposed aromatic residues, such as Phe 43 and Tyr 242 near the grooves may be important in conferring preferential binding for ssRNA.

Although the magnesium-dependent NTPase activity can be attributed to the monomeric subunit, the RNA binding, helix-destabilizing activity and possibly the interaction with the viral polymerase may require octamer formation. Integration of the two principal properties of NSP2 into one unit may have further implications. In the octamer, the proposed NTP binding sites are located on either side of the groove (open arrow in Fig. 3c). Although NTP hydrolysis is not directly linked either to the RNA binding, or to the helix-destabilizing activity of NSP2, it is possible that binding of nucleotide into the cleft alters the conformation of the monomer and affects octamer–RNA interactions. Indeed, conformational changes in the octamer have been observed upon binding of NTP and/or RNA using analytical ultracentrifugation studies⁹. Such allosteric changes may facilitate translocation of the bound RNA through the polymerase during replication, resulting in concomitant synthesis of duplex RNA and its packaging into assembling particles.

We have thus presented here the first structure of a dsRNA virus-encoded non-structural protein that is involved in genome replication and packaging. Our structural studies substantiate the proposed notion⁹ that NSP2 may function as a molecular motor during genome replication and packaging. We find that the C-terminal domain exhibits a HIT-like fold. Although this fold is observed in several cellular proteins, our studies represent the first observation of this fold in the context of viral proteins. At present there is no reason to believe that NSP2 and the cellular HIT proteins evolved from the same ancestral protein, so perhaps the similarity here suggests a convergent evolution. □

Methods

The X-ray data and coordinates of NSP2 have been deposited in the Protein Data Bank under entry code 1L9V.

Purification and expression of NSP2

Native NSP2 was expressed in *E. coli* strain M15 (pREP5) and purified as described before⁵. The protein was concentrated to 6–8 mg ml⁻¹ using a centrifugal concentrator (Ultracel PL-10, $M_r = 30,000$ cutoff) in a low-salt buffer (LSB, 2 mM Tris-HCl pH 7.2, 0.5 mM EDTA, 0.5 mM DTT). For obtaining seleno-methionine protein, cells were grown in minimal media to an absorbance at 600 nm (A_{600}) of 0.6–0.8 and induced by adding 1 mM IPTG for a duration of 14 h at 37 °C. The cell pellet from 1 L culture was resuspended in 25 ml of lysis buffer (LSB, 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 2 mg ml⁻¹ lysozyme and 0.5 mg ml⁻¹ RNase A), and incubated on ice for 2 h. Cells were disrupted by sonication with a microprobe sonicator. Addition of RNase A to the lysis buffer was necessary to obtain increased yields and solubilization of the Se-Met NSP2. The clarified lysate was bound to 2 ml of Qiagen Ni-NTA agarose beads in the batch mode and washed twice with buffer (300 mM NaCl, 10 mM Tris-HCl pH 8.0) containing 20 mM and 50 mM imidazole. The protein was eluted with 50 ml of buffer containing 200 mM imidazole, and then dialysed twice against a litre of the final crystallization buffer (CSB, 10 mM Tris-HCl pH 7.2, 100 mM NaCl, 3.5 mM TCEP, 0.5 mM EDTA) for 1–2 days. After dialysis, the protein was concentrated as described above.

Crystallization

The native protein in LSB was nearly 80% monodisperse and formed octamers as analysed by dynamic light scattering using DynaPro (Protein Solution). Addition of 200 mM MgCl₂ to the protein in LSB made it almost 95% monodisperse and allowed crystallization. Crystals were grown in all cases by the hanging drop method by equilibrating a 1:1 mixture of protein (6–8 mg ml⁻¹) with the well solution. For native protein, the well solution contained PEG 6000 (12–15%), 0.2 M magnesium acetate, 100 mM Tris-HCl pH 6.5–7.5. In the case of the seleno-methionine protein, the protein was suspended in 100 mM MgCl₂ to make it monodisperse, followed by crystallization using PEG 6K (12.5–17.5%), magnesium acetate (0.1 to 0.2 M) and HEPES/Na cacodylate (pH 6.5–7.5). In all cases, plate-like crystals appeared between 1–4 days and took two to three weeks to reach their optimum size (0.08–0.1 mm × 0.4 mm × 0.4 mm).

Data collection and processing

Crystals were suspended in mother liquor with 5% and then 10% glycerol for 1 min each and transferred into 20% glycerol. These crystals were flash frozen in liquid nitrogen before shipment to the synchrotron facility. For both native and seleno-methionine crystals, diffraction data were collected at 101 K at the National Synchrotron Light Source (Upton, New York) on beamline X8C. For seleno-methionine-NSP2 crystals, multiwavelength anomalous diffraction (MAD) data were collected at the peak of selenium K-edge (0.9789 Å) and a remote wavelength (0.9611 Å) before the crystal showed signs of radiation damage. Data processing was done using DENZO²³ and scaled with SCALEPACK²³. Crystals belonged to the I422 space group with unit cell dimensions of $a = 108.357 \text{ Å}$, $b = 108.357 \text{ Å}$, $c = 152.317 \text{ Å}$, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$.

Structure determination and refinement

The anomalous scaled MAD data were further scaled using the CCP4²⁴ program 'revise'. The subsequent processing was carried out using the CNS²⁵ suite of programs. Out of 11 seleno-methionines, 10 were located using anomalous and dispersive difference Patterson maps calculated using the diffraction data at peak and remote wavelengths. A map with FOM of 0.63 was calculated to 3.2 Å resolution. Model building was carried out manually using the program O²⁶, and several rounds of torsional refinement were performed using CNS. Rigid body refinement with this model was carried out on the native data followed by several rounds of simulated annealing refinement and model building to extend the resolution to 2.6 Å using CNS. Any model bias was examined by computing composite-OMIT maps at every stage. The final structure has an R_{free} (ref. 27) of 28.8%, calculated using 10% of the data that were not included in the refinement, and a final R -factor of 24% with the remaining 13,324 reflections, which represented ~93% completeness. In the last resolution shell (2.7 Å–2.6 Å), the data was 94.2% complete. The model (root mean square (r.m.s.) deviation in bond length, 0.007 Å, in bond angles, 1.25°) has 84% residues in the maximum allowed region, a single residue in the generously allowed region, and no residues in the disallowed region. Other crystallographic data and refinement statistics are provided in the Supplementary Information. Ribbon diagrams were computed using MOLSCRIPT²⁸ and RASTER3D²⁹ (Figs 1 and 3) or PyMol (<http://www.pymol.org>) as a renderer (Fig. 1b). Electrostatic potential map of the octamer was calculated using DelPhi, visualized using GRASP³⁰, and rendered (Fig. 3b, c) using the POV4GRASP utility (<http://pov4grasp.free.fr>).

Received 19 November 2001; accepted 25 March 2002.

- Estes, M. K. *et al.* in *Virology* (ed. Fields, B. N.) 1625–1655 (Raven, New York, 1996).
- Prasad, B. V. V. *et al.* Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* **382**, 471–473 (1996).
- Reinisch, K. M., Nibert, M. L. & Harrison, S. C. Structure of the reovirus core at 3.6-Å resolution. *Nature* **404**, 960–967 (2000).
- Grimes, J. M. *et al.* The atomic structure of the bluetongue virus core. *Nature* **395**, 470–478 (1998).
- Taraporewala, Z., Chen, D. & Patton, J. Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. *J. Virol.* **73**, 9934–9943 (1999).
- Taraporewala, Z. & Patton, J. Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. *J. Virol.* **75**, 4519–4527 (2001).
- Aponte, C., Poncet, D. & Cohen, J. Recovery and characterization of a replicase complex in rotavirus-infected cells by using a monoclonal antibody against NSP2. *J. Virol.* **70**, 985–991 (1996).
- Gallegos, C. & Patton, J. Characterization of rotavirus replication intermediates: a model for the assembly of single-shelled particles. *Virology* **172**, 616–627 (1989).
- Schuck, P., Taraporewala, Z., McPhie, P. & Patton, J. T. Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes. *J. Biol. Chem.* **276**,

9679–9687 (2001).

- Lima, C. D., Klein, M. G. & Hendrickson, W. A. Structure-based analysis of catalysis and substrate definition in the HIT protein family. *Science* **278**, 286–290 (1997).
- Petrie, B. L., Greenberg, H. B., Graham, D. Y. & Estes, M. K. Ultrastructural localization of rotavirus antigens using colloidal gold. *Virus Res.* **1**, 133–152 (1984).
- Kattoura, M., Chen, X. & Patton, J. The rotavirus RNA-binding protein NS35 (NSP2) forms 10S multimers and interacts with the viral RNA polymerase. *Virology* **202**, 803–813 (1994).
- Ramig, R. & Petrie, B. L. Characterization of temperature-sensitive mutants of simian rotavirus SA11: protein synthesis and morphogenesis. *J. Virol.* **49**, 665–673 (1984).
- Chen, D., Gombold, J. L. & Ramig, R. F. Intracellular RNA synthesis directed by temperature-sensitive mutants of simian rotavirus SA11. *Virology* **178**, 143–151 (1990).
- Uitenweerde, J. M., Theron, J., Stoltz, M. A. & Huismans, H. The multimeric nonstructural NS2 proteins of bluetongue virus, African horsesickness virus, and epizootic hemorrhagic disease virus differ in their single-stranded RNA-binding ability. *Virology* **209**, 624–632 (1995).
- Taraporewala, Z., Chen, D. & Patton, J. Multimers of the bluetongue virus nonstructural protein, NS2, possess nucleotidyl phosphatase activity: similarities between NS2 and rotavirus NSP2. *Virology* **280**, 221–231 (2001).
- Gillian, A. L., Schmechel, S. C., Livny, J., Schiff, L. A. & Nibert, M. L. Reovirus protein sigma NS binds in multiple copies to single-stranded RNA and shares properties with single-stranded DNA binding proteins. *J. Virol.* **74**, 5939–5948 (2000).
- Cohen, J. Rethinking a vaccine risk. *Science* **293**, 1576–1577 (2001).
- Holm, L. & Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acid Res.* **27**, 244–247 (1999).
- Lima, C. D., Klein, M. G., Weinstein, I. B. & Hendrickson, W. A. Three-dimensional structure of human protein kinase C interacting protein 1, a member of the HIT family of proteins. *Proc. Natl. Acad. Sci. USA* **28**, 5357–5362 (1996).
- Brenner, C., Biegonowski, P., Pace, H. C. & Huebner, K. The histidine superfamily of nucleotide-binding proteins. *J. Cell. Phys.* **181**, 179–187 (1999).
- Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Struct. Biol.* **7**, 648–652 (2000).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank A. Nickitenko, B. Bowman, W. Meador and the National Synchrotron Light Source beamline (X8C) staff for help with data collection, and F.A. Quiocho for allowing us to use X-ray diffraction facilities at the Baylor College of Medicine. We also thank M. K. Estes and R. F. Ramig, and M. Sowa, for helpful discussions. This work was supported by grants from the NIH and R. Welch foundation to B.V.V.P. and an NIH intramural grant to J.P.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.V.V.P. (e-mail: vprasad@bcm.tmc.edu).

Microbiology media and methods

Detection, growth and sterilization.

Disposable flasks

Nalgene www.nalgenunc.com

Reduce the risk of cross-contamination

Nalgene's sterile disposable flasks are made of clear PETG plastic and come with a plain or baffled base, making them suitable for bench-top or shaker table use. The flasks have graduations for volume assessment and are provided with a leak-proof polyethylene screw closure that can be vented with a quarter turn. The flasks are in the Erlenmeyer design and are available in 250, 500 and 2,000 ml capacities.



A throw-away line from Nalgene.

Vitek 2

bioMérieux www.biomérieux.com

Ask an expert

The Vitek 2 Advanced Expert system (AES) is a tool for the rapid detection and interpretation of key resistance mechanisms. It is designed to enable microbiologists to gather the information on which to base therapeutic strategies. The system features automated susceptibility testing for rapid determination of minimum inhibitory concentrations by analysing the growth kinetics of bacteria in the presence of different antibiotic concentrations. A database containing over 20,000 MICs and describing more than 2,000 phenotypes is used to validate and interpret the results of susceptibility tests.

GIBCO OptiPro

Invitrogen www.invitrogen.com

Versatile serum-free medium

OptiPro SFM for virus and recombinant protein production contains no components of animal or human origin, reducing the risk of virus or prion contamination. GIBCO claims that cell growth and virus production are equivalent or superior to those obtained using serum-supplemented systems, and points out that this product's low protein con-

centration ($7.5 \mu\text{g ml}^{-1}$) facilitates downstream purification. OptiPro sustains the growth of a range of kidney epithelial cell lines and can produce multiple viruses to high titre in multiple kidney-derived cell lines.

ProtoCOL

Synbiosis www.synbiosis.com

Automated colony counter

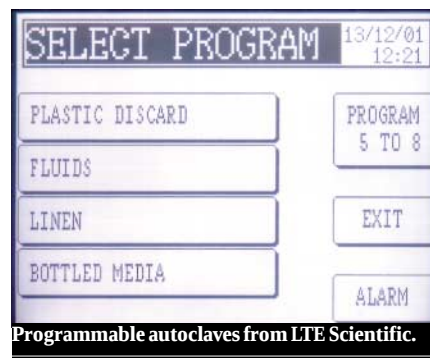
The ProtoCOL automated spiral plate counter generates precise counts of propionibacteria, staphylococci and the yeast, *Malassezia furfur*, from a variety of selective media. It also records details of the sample, including a plate image, data and time, in a printable format. It is especially suitable for use in clinical trials of anti-infective therapies.

LB Agars and Broths

AMRESCO www.amresco-inc.com

Something in the media

LB Agars and Broths for Miller and Lennox formulations are supplied as powdered blends of media and agar conveniently packaged in single-use foil pouches. When the contents of a pouch are dissolved in water and autoclaved, it prepares 1 litre of medium. LB Agars and Broths boast long stability and shelf life and come in convenient pre-measured blends.



Programmable autoclaves from LTE Scientific.

Touchclave

LTE Scientific www.lte-scientific.co.uk

Big and small, autoclaves for all

Touchclave autoclaves are available in sizes from 300 to 7,000 litres. The sterilization cycle features a two-stage air removal phase to conserve power and water. Sub-atmospheric steam flushing and positive pulsation take place prior to steam sterilization, followed by a deep-drying vacuum. Standard features include an integral printer and independent recorder and a diagnostic system to highlight abnormalities in the cycle.

BIO 101 Systems

Qbiogene www.qbiogene.com

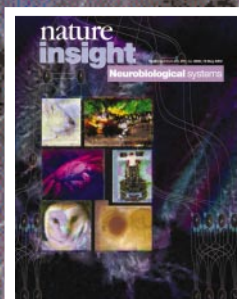
Say yes to this line of growth media

This line of growth media includes pre-mixed media for *S. cerevisiae* and *S. pombe* as well as general formulations of YPD/YEPD, YEP, YPAD, YE, YES and ME mixtures. Synthetic and minimal defined media formulations and custom formulations for specialized experiments are also available. The products can be supplied as media capsules, pre-mixed powders in pouches, bottles or drums, liquid media and pre-poured plates.

These notes are compiled in the Nature office from information provided by the manufacturers.

nature insight

Neurobiological systems



Cover illustration

Each of the experimental preparations described in this Insight — birds, flies, crabs, salamanders and even robots — has unique aspects that contribute to our understanding of general neural principles.

In an age of powerful genetic tools to study mice and men, one might wonder why neuroscientists would ever stray from mammalian systems. The molecules-to-behaviour approach has offered important insights into areas of brain research including learning and memory, and disease. But, as Charles Sherrington pointed out almost a century ago in his classic lectures on the *Integrated Action of the Nervous System*, it is in its functional interconnectivity that the study of the nervous system assumes its due importance. The human brain contains several billion neurons, and several trillion interconnections. On the assumption that there are some common principles of nervous system action — which seems a reasonable proposition from an evolutionary standpoint — attempting to translate circuits to behaviour on the massive human-brain scale poses some significant drawbacks.

Thus, one might argue that the ideal place to begin an integrative understanding of a neural system is one in which the behaviour is particularly well defined and in which the neural circuitry is accessible and can be equally well defined. The reviews in this Insight are meant to sample the advantages of several experimental preparations in elucidating general neural principles. As Eve Marder points out in her overview, there were many more such examples left out than included. Those here span a neuroscientific range from circadian rhythms in *Drosophila*, for which the molecular backbone is in place to begin teasing apart systems-level questions, to robots based on animal design, created to test our understanding of circuits that are 'solved.' They provide an interesting and varied sample of some of the very elegant neuroscience that is underway outside the mammalian world, and yet affects our understanding of mammalian neuroscience.

We are pleased to acknowledge the financial support of the National Institutes of Health in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer review.

Hemai Parthasarathy Senior Editor

overview:

- 318 Non-mammalian models for studying neural development and function**
E. Marder

review articles:

- 322 Instructed learning in the auditory localization pathway of the barn owl**
E. I. Knudsen

- 329 Circadian rhythms from flies to human**
S. Panda, J. B. Hogenesch & S. A. Kay

- 336 On the scents of smell in the salamander**
J. S. Kauer

- 343 A small-systems approach to motor pattern generation**
M. P. Nusbaum & M. P. Beenhakker

- 351 What songbirds teach us about learning**
M. S. Brainard & A. J. Doupe

- 359 Robots in invertebrate neuroscience**
B. Webb

Editor, *Nature*: Philip Campbell
Insights Editor: Lesley Anson
Editorial Assistant: Simon Gibson
Production Editor: Simon Gribbin

Art Director: Majo Xeridat
Diagrams: Ann Thomson
Vicky Askew
Suzanne Coleman

Production Manager: Yvonne Strong
Sponsorship: Emma Jones

Non-mammalian models for studying neural development and function

Eve Marder

Volen Center, MS 013, Brandeis University, Waltham, Massachusetts 02454-9110, USA (e-mail: marder@brandeis.edu)

Early neuroscientists scoured the animal kingdom for the ideal preparation with which to study specific problems of interest. Today, non-mammalian nervous systems continue to provide ideal platforms for the study of fundamental problems in neuroscience. Indeed, the peculiarities of body plan and nervous systems that have evolved to carry out precise tasks in unique ecological niches enable investigators not only to pose specific scientific questions, but also to uncover principles that are general to all nervous systems.

Not that long ago, neuroscience graduate students were expected to wander the woods, explore tide pools, take ocean voyages, or pore over tomes of zoological texts with wonderful old drawings in search of the perfect preparation with which to study an important problem. In this they were following the lead of their elders. Many of the heroic figures among early neuroscientists avidly sought through the animal kingdom for the ideal preparation with which to study the problem that interested them, and some, like Ted Bulloch and Steve Kuffler, studied many different preparations during their careers. Furshpan and Potter¹ first studied electrical coupling in crayfish, Kuffler, Nicholls and Orkand first recorded intracellularly from *Necturus* (an amphibian) glial cells^{2,3}, Hodgkin and Huxley used the squid giant axon to understand the mechanism of the action potential⁴, Dudel and Kuffler⁵ first used quantal analysis to demonstrate presynaptic inhibition at the crustacean neuromuscular junction, and Ratliff and Hartline first described lateral inhibition in photoreceptors of the horseshoe crab *Limulus*⁶. Levi-Montalcini and Viktor Hamburger did seminal work on the development of the nervous system using chick embryos⁷⁻⁹, and frogs and fish were the early preparations of choice for the study of the specificity of retinal-tectal projections in development¹⁰⁻¹². Retinal structure and function was studied in fish, salamander and turtle retinas¹³⁻¹⁶, and birds, bats and electric fish were favoured for studies of other sensory modalities¹⁷⁻²¹.

Journey's end

Today it is almost inconceivable that many neuroscientists would venture back to the ocean, river, field or forest in search of a new preparation. This is for several reasons — practical, philosophical and political. First, and perhaps most important, we now have developed large collections of data on a number of systems upon which new studies build. Initiating studies on the nervous system of an animal on which there is no literature would require a forbidding



amount of groundwork to bring it to the level of one of the more established preparations. Second, the pressure for direct medical relevance has pushed neuroscience towards studies of animals thought to be good models of human function and dysfunction. Third, the availability of genetic tools in some organisms has significantly enhanced their power, and thus attractiveness. And fourth, scientists, like other humans, are too often conformists.

That said, there are some who have even recently developed new non-mammalian preparations or turned

to existing preparations to address questions not previously studied with these animals. For example, Ron Hoy and colleagues have continued to go to the field to find organisms with fascinating attributes, focusing on relatively unstudied insects and jumping spiders for their unusual sensory organs and astonishing behaviour²²⁻²⁶. This work reminds us that the study of neurobiological mechanisms in the context of their natural setting, as is the goal of neuroethology, brings us closest to the fundamental lessons of evolution and natural species diversity.

One of the most exciting areas in systems neuroscience is the cellular and circuit mechanisms underlying sustained neural activity and its role in working memory. Much of the work that has defined these issues has been done in awake and behaving monkeys^{27,28}. David Tank and Sebastian Seung have recently started studying these issues using the vestibular ocular reflex of fish²⁹⁻³². The fish learn, it is relatively easy to combine behavioural and electrophysiological measurements in them, and they allow a variety of mechanistic experiments not possible or practical with primates.

Of course, there are many who continue to exploit non-mammalian preparations to study a raft of important problems in neuroscience, only a few of which could be highlighted in this collection of reviews. Although the era that saw the proliferation of preparations has ended, scores of animals, from worms to birds, continue to instruct us. Of these, several have been selected that illustrate how non-mammalian preparations are today catalysing discovery in neuroscience. It would have been just as easy to select



The peculiarities of body plan and nervous system that have evolved as animals colonized strange ecological niches offer huge potential for deducing general principles that will be applicable to all nervous systems.

wonderful work on bat echolocation^{33,34}, zebrafish development and behaviour^{35,36}, electric-field sensing and generation in electric fish and eels^{37,38}, insect sensory processing^{22-26,39-41}, leech, tadpole and lamprey swimming^{42,43}, behaviour of the nematode worm *Caenorhabditis elegans*^{44,45} or sexual dimorphism in the frog nervous system associated with courtship behaviours⁴⁶, to name only a few. In all of these preparations, the peculiarities of these animals have allowed investigators to pose specific scientific questions into basic mechanisms of sensory-motor integration and their relation to behaviour. In each of the articles that follow, the authors have attempted to bring the reader a sense of how each preparation has led to better understanding of a basic question in neuroscience.

Nervous systems, learning and behaviour

Learning is required for animals to adapt successfully both to their environment and to changes in their own body. We recently saw the Nobel prize awarded to Eric Kandel⁴⁷ for his work on the cellular basis of learning using the sea slug, *Aplysia californica*. Kandel's choice of this mollusc, with its orange-coloured 'simple' nervous system, was crucial in the early attempts to tackle the formidable task of uncovering the cellular and molecular mechanisms underlying simple forms of learning. The small size of the animal's nervous system, the simplicity of its behaviours, and the ability to easily identify *Aplysia* neurons facilitated attempts aimed at determining the sites at which changes during learning might occur. Kandel and his colleagues have made extraordinary progress identifying the cellular and molecular mechanisms by which alterations in synaptic strength are produced by a variety of stimulus paradigms at some of the loci in the animal that are likely to be responsible for stable changes in behaviour.

Certainly, much remains to be understood about how learning in *Aplysia* takes place. But in this system one can imagine ultimately discovering answers to questions such as does behavioural learning involve changes at most of the synapses in a set of pathways, how distributed are the changes in the nervous system underlying behavioural modifications, and how do all the changes that occur during learning work together to produce an altered or modified behaviour? These are questions that are crucial for understanding learning in all nervous systems, but are difficult to study in mammalian brains because of the large number of neurons and connections.

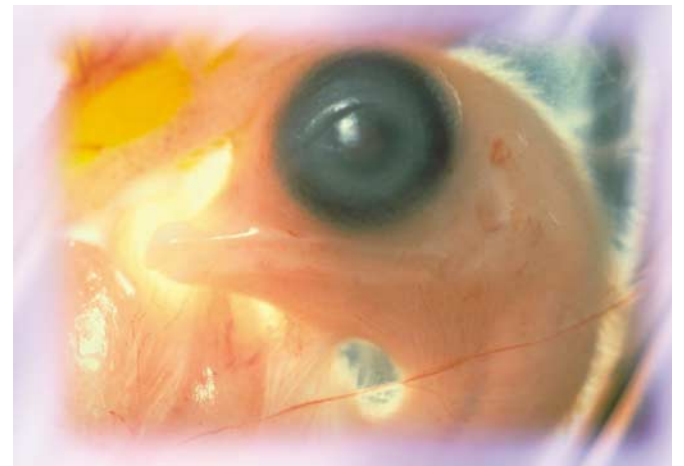
Learning is studied in organisms throughout the animal kingdom from *C. elegans* to humans. 'Birdbrain' might be a common colloquialism used to insult a person's intelligence, but bird brains provide outstanding opportunities to study higher cognitive function in remarkable ways. Two articles in this issue discuss learning in auditory processes of birds in the context of two very different



Tremendous progress has been made in identifying the cellular and molecular mechanisms that underlie simple forms of behavioural learning in the sea slug *Aplysia californica*.

problems — sensory map formation and song production. Barn owls localize sound exceptionally well, and use this ability to hunt for their prey even in limited light. Knudsen (pages 322–328) discusses the structure of the owl's auditory space map, how it develops and how it is modified by experience. The auditory space map is adjusted by visual experience, and the problem of how different sensory maps are brought into register by experience is one that is beautifully posed and studied in owls. Recent work described by Knudsen seeks to define the loci for change in the brain responsible for behavioural change and then to bring these changes down to cellular and molecular mechanisms.

Brainard and Doupe (pages 351–358) focus on learning in the birdsong system, an area that provides one of the richest sets of questions in neuroscience and neuroethology. Here it is possible to directly address questions such as how a complex motor behaviour is learned, how complex auditory sequences are decoded, how sexually dimorphic brain structures are controlled by hormones during development, and what cellular and molecular changes underlie 'critical periods', the times at which critical experiences are crucial for the appropriate development of the nervous system to occur. As in the barn owl system, work in the songbird system is anchored in behavioural studies showing the animal's capacity for sensory and motor performance and learning. These observations have then been



Seminal work by Viktor Hamburger and co-workers established the chick embryo as an animal model for experimental embryology and marked the beginning of the new discipline of developmental neuroscience.



Few students of the biological sciences will have graduated without encountering the classical studies of Hodgkin and Huxley, who investigated the mechanism of the action potential using the squid giant axon.

complemented by forays into the brain to discover the neural circuits and cellular processes that produce the behaviours and their plasticity. It is precisely this solid neuroethological anchor that has made these preparations so instructive for understanding how brains produce complex behaviours.

Conservation in construction

Olfaction is central to many animals as they find food and mates. To the surprise of many, the organization of the olfactory system both at the molecular and circuit level is remarkably conserved across species from worms, molluscs, insects, salamanders and rodents^{48–52}. Common themes have emerged from the study of the roles of oscillations in odour processing, and this is a field in which invertebrate, non-mammalian vertebrate (see review by Kauer, pages 336–342) and rodent work continues to inform.

The commonality of mechanism across phylogenetic boundaries is illustrated in the review by Panda *et al.* (pages 329–335) on circadian rhythms. Although the existence of circadian rhythms has been long known, it was the discovery of rhythm mutants in the fruitfly *Drosophila melanogaster*^{53,54} that ushered in the modern era of circadian rhythm research. Using *Drosophila*, a several laboratories have isolated a number of genes that are part of the circadian clock⁵⁵, and these have led to models of molecular and biochemical feedback loops that can account for circadian rhythmicity. Many of the molecular components of the clock that were first described in flies have since been found in mammals. This is a prime example of a discovery process that depends heavily on the ease of doing genetics and behaviour in an organism that develops quickly and in which thousands of lines can be rapidly screened.

Almost twenty years ago, the analysis of small invertebrate motor systems triggered a paradigm shift in our thinking of how networks generate behaviour^{56–58}. Before that time, it was generally believed that networks were static, and that it would be sufficient to work out the 'wiring diagram' or 'connectivity diagram' to understand how a given network operates. But work on small motor systems showed that networks can be reconfigured to produce multiple outputs, as the synaptic strengths and intrinsic properties of network neurons are modified by synaptic inputs and neuromodulatory substances⁵⁹. Nusbaum and Beenhakker (pages 343–350) describe work on one of the canonical small motor systems, the stomatogastric ganglion of lobsters and crabs. This work illustrates that networks are modulated by multiple inputs, and provides direct examples of the kinds of

mechanisms that can be used to tune the same network to produce a variety of circuit dynamics.

There is today an uneasy flirtation between the fields of neuroscience and artificial intelligence and robotics. A growing number of investigators look to invertebrate nervous systems for design principles in the construction of robots that can sense their environment and navigate intelligently through it. Webb (pages 359–363) describes much of this work, and also argues that the construction of robots based in what is known about some 'simple' invertebrates can also help neuroscientists understand the limitations of their knowledge about these preparations.

General principles from the arcane

For good or ill, some of the preparations that in the past were exploited for significant scientific gain have fallen by the wayside, and have become historical oddities. But to concentrate all resources into the collective study of a very few nervous systems would be a pity on both practical and philosophical grounds. Those who sit on government advisory panels and urge that all funding for neuroscience be used to support mouse, monkey and human work forget the interdependence of species in the survival of our planet. They forget our wonder as we spot an unusual bird in the mangroves of Florida or the jungles of Malaysia. As we revel in the sometimes outrageous forms that species take, we remember that species diversity was an outcome of survival in disparate environments. The peculiarities of body plan and nervous systems that allowed animals to live in strange ecological niches remind me that the most important findings in science often result from individual scientists' foibles, genius, insight and personal taste. Although brute-force science has its place, we risk an incalculable loss of individual creativity and imagination if we work only on consensus problems and consensus preparations. We should value and protect those who dare to be fascinated by animals that have evolved nervous systems to best carry out a specific task.

New technologies are expanding the range of approaches available in the study of all nervous systems. By studying the neural mechanisms underlying the processes in the animals ideally suited for their analysis we stand the best chance of finding the principles that will be general to all nervous systems, including our own. □

1. Furshpan, E. J. & Potter, D. D. Transmission at the giant motor synapses of the crayfish. *J. Physiol.* **145**, 289–325 (1959).
2. Kuffler, S. W., Nicholls, J. G. & Orkand, R. K. Physiological properties of glial cells in the central nervous system of amphibia. *J. Neurophysiol.* **29**, 768–787 (1966).
3. Kuffler, S. W. Neuroglial cells: physiological properties and a potassium mediated effect of neuronal activity on the glial membrane potential. *Proc. R. Soc. Lond. B* **168**, 1–21 (1967).
4. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544 (1952).
5. Dudel, J. & Kuffler, S. W. Presynaptic inhibition at the crayfish neuromuscular junction. *J. Physiol.* **155**, 543–562 (1961).
6. Lange, D., Hartline, H. K. & Ratliff, F. Inhibitory interaction in the retina: techniques of experimental and theoretical analysis. *Ann. NY Acad. Sci.* **128**, 955–971 (1966).
7. Hamburger, V. Ontogeny of neuroembryology. *J. Neurosci.* **8**, 3535–3540 (1988).
8. Hamburger, V. History of the discovery of neuronal death in embryos. *J. Neurobiol.* **23**, 1116–1123 (1992).
9. Hamburger, V. The history of the discovery of the nerve growth factor. *J. Neurobiol.* **24**, 893–897 (1993).
10. Gaze, R. M. & Keating, M. J. The depth distribution of visual units in the tectum of the frog following regeneration of the optic nerve. *J. Physiol.* **200**, 128P–129P (1969).
11. Gaze, R. M. & Sharma, S. C. Axial differences in the reinnervation of the goldfish optic tectum by regenerating optic nerve fibres. *Exp. Brain Res.* **10**, 171–181 (1970).
12. Jacobson, M. & Gaze, R. M. Selection of appropriate tectal connections by regenerating optic nerve fibers in adult goldfish. *Exp. Neurol.* **13**, 418–430 (1965).
13. Baylor, D. A. & Fuortes, M. G. Electrical responses of single cones in the retina of the turtle. *J. Physiol.* **207**, 77–92 (1970).
14. Baylor, D. A., Fuortes, M. G. & O'Bryan, P. M. Receptive fields of cones in the retina of the turtle. *J. Physiol.* **214**, 265–294 (1971).
15. Baylor, D. A. & Lamb, T. D. Local effects of bleaching in retinal rods of the toad. *J. Physiol.* **328**, 49–71 (1982).
16. Baylor, D. A., Matthews, G. & Nunn, B. J. Location and function of voltage-sensitive conductances in retinal rods of the salamander, *Ambystoma tigrinum*. *J. Physiol.* **354**, 203–223 (1984).
17. Carr, C. E. & Konishi, M. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* **10**, 3227–3246 (1990).
18. Carr, C. E. & Konishi, M. Axonal delay lines for time measurement in the owl's brainstem. *Proc. Natl Acad. Sci. USA* **85**, 8311–8315 (1988).
19. Volman, S. F. & Konishi, M. Spatial selectivity and binaural responses in the inferior colliculus of the great horned owl. *J. Neurosci.* **9**, 3083–3096 (1989).

20. Carr, C. E. Time coding in electric fish and barn owls. *Brain Behav. Evol.* **28**, 122–133 (1986).
21. Konishi, M. The neural algorithm for sound localization in the owl. *Harvey Lect.* **86**, 47–64 (1990).
22. Faure, P. A. & Hoy, R. R. The sounds of silence: cessation of singing and song pausing are ultrasound-induced acoustic startle behaviors in the katydid *Neoconocephalus ensiger* (Orthoptera; Tettigoniidae). *J. Comp. Physiol. A* **186**, 129–142 (2000).
23. Buschbeck, E., Ehmer, B. & Hoy, R. Chunk versus point sampling: visual imaging in a small insect. *Science* **286**, 1178–1180 (1999).
24. Crocroft, R. B., Tieu, T. D., Hoy, R. R. & Miles, R. N. Directionality in the mechanical response to substrate vibration in a treehopper (Hemiptera: Membracidae: *Umbonia crassicornis*). *J. Comp. Physiol. A* **186**, 695–705 (2000).
25. Mason, A. C., Oshinsky, M. L. & Hoy, R. R. Hyperacute directional hearing in a microscale auditory system. *Nature* **410**, 686–690 (2001).
26. Farris, H. E. & Hoy, R. R. Two-tone suppression in the cricket, *Eunemobius carolinus* (Gryllidae, Nemobiinae). *J. Acoust. Soc. Am.* **111**, 1475–1485 (2002).
27. Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
28. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X. J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
29. Seung, H. S., Lee, D. D., Reis, B. Y. & Tank, D. W. The autapse: a simple illustration of short-term analog memory storage by tuned synaptic feedback. *J. Comput. Neurosci.* **9**, 171–185 (2000).
30. Aksay, E., Gamkrelidze, G., Seung, H. S., Baker, R. & Tank, D. W. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature Neurosci.* **4**, 184–193 (2001).
31. Aksay, E., Baker, R., Seung, H. S. & Tank, D. W. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *J. Neurophysiol.* **84**, 1035–1049 (2000).
32. Seung, H. S., Lee, D. D., Reis, B. Y. & Tank, D. W. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* **26**, 259–271 (2000).
33. Zhang, Y. & Suga, N. Corticofugal amplification of subcortical responses to single tone stimuli in the mustached bat. *J. Neurophysiol.* **78**, 3489–3492 (1997).
34. Gao, E. & Suga, N. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proc. Natl Acad. Sci. USA* **95**, 12663–12670 (1998).
35. Lorent, K., Liu, K. S., Fetcho, J. R. & Granato, M. The zebrafish space cadet gene controls axonal pathfinding of neurons that modulate fast turning movements. *Development* **128**, 2131–2142 (2001).
36. Ritter, D. A., Bhatt, D. H. & Fetcho, J. R. In vivo imaging of zebrafish reveals differences in the spinal networks for escape and swimming movements. *J. Neurosci.* **21**, 8956–8965 (2001).
37. Kreiman, G., Krahe, R., Metzner, W., Koch, C. & Gabbiani, F. Robustness and variability of neuronal coding by amplitude-sensitive afferents in the weakly electric fish eigenmannia. *J. Neurophysiol.* **84**, 189–204 (2000).
38. Lewis, J. E. & Maler, L. Neuronal population codes and the perception of object distance in weakly electric fish. *J. Neurosci.* **21**, 2842–2850 (2001).
39. Gabbiani, F., Mo, C. & Laurent, G. Invariance of angular threshold computation in a wide-field looming-sensitive neuron. *J. Neurosci.* **21**, 314–329 (2001).
40. Paydar, S., Doan, C. A. & Jacobs, G. A. Neural mapping of direction and frequency in the cricket cercal sensory system. *J. Neurosci.* **19**, 1771–1781 (1999).
41. Jacobs, G. A. & Theunissen, F. E. Extraction of sensory parameters from a neural map by primary sensory interneurons. *J. Neurosci.* **20**, 2934–2943 (2000).
42. Calabrese, R. L. Cellular, synaptic, network, and modulatory mechanisms involved in rhythm generation. *Curr. Opin. Neurobiol.* **8**, 710–717 (1998).
43. Grillner, S., Wallen, P., Brodin, L. & Lansner, A. Neuronal network generating locomotor behavior in lamprey: circuitry, transmitters, membrane properties, and simulation. *Annu. Rev. Neurosci.* **14**, 169–199 (1991).
44. Pierce-Shimomura, J. T., Morse, T. M. & Lockery, S. R. The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J. Neurosci.* **19**, 9557–9569 (1999).
45. Rose, J. K. & Rankin, C. H. Analyses of habituation in *Caenorhabditis elegans*. *Learn. Mem.* **8**, 63–69 (2001).
46. Tobias, M. L., Viswanathan, S. S. & Kelley, D. B. Rapping, a female receptive call, initiates male-female duets in the South African clawed frog. *Proc. Natl Acad. Sci. USA* **95**, 1870–1875 (1998).
47. Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
48. Sengupta, P., Chou, J. H. & Bargmann, C. I. odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84**, 899–909 (1996).
49. Sengupta, P., Colbert, H. A. & Bargmann, C. I. The *C. elegans* gene odr-7 encodes an olfactory-specific member of the nuclear receptor superfamily. *Cell* **79**, 971–980 (1994).
50. Gelperin, A. Oscillatory dynamics and information processing in olfactory systems. *J. Exp. Biol.* **202**, 1855–1864 (1999).
51. Ermentrout, B., Wang, J. W., Flores, J. & Gelperin, A. Model for olfactory discrimination and learning in *Limax* procerebrum incorporating oscillatory dynamics and wave propagation. *J. Neurophysiol.* **85**, 1444–1452 (2001).
52. Cooke, I. R. & Gelperin, A. In vivo recordings of spontaneous and odor-modulated dynamics in the *Limax* olfactory lobe. *J. Neurobiol.* **46**, 126–141 (2001).
53. Konopka, R. J. & Benzer, S. Clock mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
54. Konopka, R. J. Genetics of biological rhythms in *Drosophila*. *Annu. Rev. Genet.* **21**, 227–236 (1987).
55. Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–119 (2001).
56. Getting, P. A. & Dekin, M. S. Mechanisms of pattern generation underlying swimming in *Tritonia*. IV. Gating of central pattern generator. *J. Neurophysiol.* **53**, 466–480 (1985).
57. Getting, P. A. Emerging principles governing the operation of neural networks. *Annu. Rev. Neurosci.* **12**, 185–204 (1989).
58. Marder, E. & Hooper, S. L. in *Model Neural Networks and Behavior* (ed. Selverston, A. I.) 319–337 (Plenum, New York, 1985).
59. Harris-Warrick, R. M. & Marder, E. Modulation of neural networks for behavior. *Annu. Rev. Neurosci.* **14**, 39–57 (1991).

Instructed learning in the auditory localization pathway of the barn owl

Eric I. Knudsen

Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA (e-mail: eknudsen@stanford.edu)

A bird sings and you turn to look at it — a process so automatic it seems simple. But is it? Our ability to localize the source of a sound relies on complex neural computations that translate auditory localization cues into representations of space. In barn owls, the visual system is important in teaching the auditory system how to translate cues. This example of instructed plasticity is highly quantifiable and demonstrates mechanisms and principles of learning that may be used widely throughout the central nervous system.

Our sense of auditory space derives from the associations we have formed between specific auditory cues and locations in the world that produce them. The dominant auditory localization cues are the relative timing and level of the sound at both ears. The central auditory system analyses these and other cues and associates particular values of these cues with locations in space. Although many animals can localize sounds soon after birth^{1,2}, the exact relationships between cue values and locations in space are shaped and modified by experience^{3–5}.

Because experience can alter sound localization, and because the pathways that mediate sound localization have been identified to a large extent, the auditory localization pathway has become a model system for studying mechanisms by which the nervous system learns from experience^{5–7}. The most extensively studied species in this regard is the barn owl (*Tyto alba*), a nocturnal predator with a highly evolved capacity for sound localization which rivals that of humans^{8,9}.

In barn owls, the pathways that process sound localization cues have been elaborated extensively, the tuning of neurons to localization cues is sharp, and the representation of auditory cue values in the brain is systematic¹⁰. These properties have enabled sensitive, quantitative assessment of experience-dependent plasticity in this species. Experimental manipulation of the owl's sensory experience has revealed functional, anatomical and pharmacological changes in the central auditory system that accompany behavioural learning. The results show an inherent advantage of innate neuronal connections over connections that are acquired with learning, a decline in learning with age, and an increased capacity for learning in adults that have had appropriate experience as juveniles. This review summarizes these findings and the mechanisms and principles that have been illuminated by them.

Experience shapes auditory orienting behaviour

Sound localization cues result from the interaction of the head and ears with the incoming sound stimulus^{9,11}. These



cues consist of interaural timing differences (ITDs), interaural level differences (ILDs) and the amplitude spectrum of the sound at each ear. ITD results from a difference in the distance that sound must travel to reach the near versus the far ear (Fig. 1a) and is the primary cue for the horizontal (azimuthal) location of a sound source (Fig. 1b). ILD and amplitude spectrum result from the frequency-dependent directional

properties of the head and ears. ILD, like ITD, varies with the azimuth of a stimulus, except in nocturnal owls such as the barn owl (Fig. 1c) for which ILD varies also with the elevation of a stimulus, owing to an asymmetry of the external ears. Amplitude spectrum has a complicated relationship with the horizontal and vertical locations of a sound source and contributes to localization in both azimuth and elevation.

To localize sound, the central nervous system (CNS) must measure the values of these cues and then associate particular cue values with the location in space that produces them. This task is complicated by the variation in the correspondence of cue values with locations in space across sound frequencies and across individuals, owing to differences in the size and shape of the head and ears. In addition, the neural representation of acoustic cues can change over the lifetime of an animal as a result of hearing loss and the development and ageing of the nervous system¹².

It is not surprising, therefore, that the auditory system calibrates its interpretation of localization cues based on experience^{3–6}. Adaptive adjustment of sound localization by barn owls has been demonstrated by subjecting owls to a variety of sensory manipulations and measuring the effects of those manipulations on the accuracy of auditory orienting behaviour, which is extremely precise in barn owls⁸. One class of manipulations has involved altering the correspondence of auditory cue values with locations in space by plugging one ear^{13,14}. Initially, monaurally occluded owls, like monaurally occluded humans, mislocalize sounds towards the side of the open ear. But after many weeks of experience with an earplug, young owls recover accurate orienting responses despite the presence of the

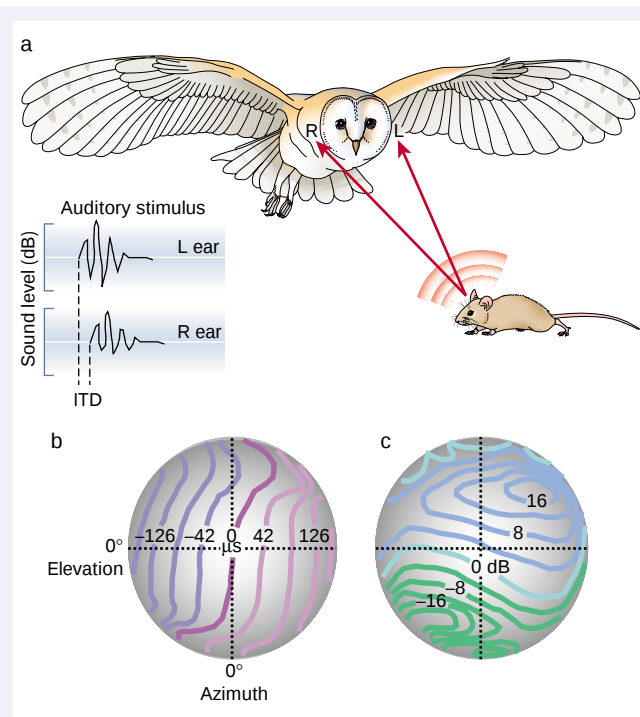


Figure 1 The relationship between auditory cue values and locations in space for a barn owl. **a**, Sound waves generated by movements of a mouse are received by the owl's left and right ears. The sound waveform in the right ear (inset) is delayed and attenuated relative to that in the left ear. **b**, **c**, Correspondence of interaural timing difference (ITD, **b**) and interaural level difference (ILD, **c**) values with locations in space for 6-kHz sound. The globes represent space around the head relative to the line of sight (the eyes of a barn owl are nearly stationary). Contour lines indicate locations that produce equivalent values of each cue. For ITD values, purple indicates left ear leading and pink indicates right ear leading, while for ILD values, green indicates left ear greater and blue indicates right ear greater. The spatial patterns of ITD and ILD change with sound frequency. The strong dependence on elevation of the ILD cue results from a vertical asymmetry of the barn owl's ears. For animals with symmetrical ears, and for barn owls at frequencies below 4 kHz, the spatial patterns for ILD are approximately symmetrical about the mid-sagittal plane, as they are for ITD. Data are from ref. 23 and are based on probe-tube measurements from the external ear canals with sounds presented in a free field.

earplug. When the earplug is removed, the owls initially make orienting errors in the opposite direction, but these errors disappear gradually with normal experience. Thus, certain manipulations of hearing cause owls to learn new associations between auditory cue values and locations in space.

A second class of sensory manipulations has involved leaving auditory cues normal, but changing the locations in the visual field to which cue values correspond by exposing owls to a visual field that is displaced by prismatic spectacles¹⁵. Because owls cannot move their eyes by more than a few degrees, those wearing prisms must learn new associations between cue values and locations in the visual field in order to bring their auditory and visual worlds into mutual alignment. This is what young owls do. Over a period of many weeks, they adjust their auditory orienting responses according to the optical displacement imposed by the prisms (Fig. 2). This learning is adaptive because it causes these owls to orient to sounds so that they see the source of the sound through the prisms.

Prism experience also causes changes in other visually guided behaviours. Just as humans adjust reaching and throwing movements when wearing prisms¹⁶, owls adjust flight and strike behaviours¹⁷. Visuomotor adjustment occurs more rapidly than auditory–visual realignment and does not decline with age. Although the plasticity that underlies visuomotor adjustment is

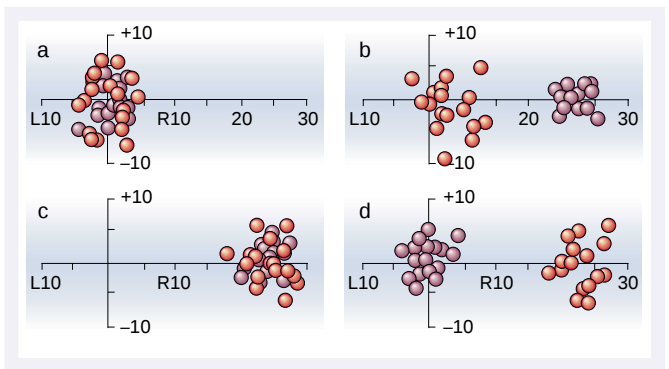


Figure 2 Plasticity of auditory orienting behaviour of a juvenile owl, resulting from prism experience. **a**, Before prisms; **b**, day 1 with visual field displaced 23° to the right by prismatic spectacles; **c**, day 42 with 23° prisms; **d**, prisms removed. Data points indicate final head orientation to auditory (red) or visual (purple) stimuli, presented in the dark, as measured with a search coil. Stimulus position was varied randomly. Responses are plotted relative to the true location of the stimulus source. Data are from ref. 72.

distinct¹⁸, the behavioural changes that result from these adjustments could help in driving auditory–visual realignment.

Neural correlates of learning

The auditory localization pathway

Neural correlates of behavioural learning are apparent in the auditory localization pathway (Fig. 3). The processing of auditory localization cues in the brainstem has been studied in a variety of species, and intensively in the barn owl. In all species, ITDs and ILDs are processed in parallel pathways and, because ITD and ILD depend on frequency, these cues are measured in frequency-specific channels^{10,19,20}. Neurons that encode these frequency-specific cue values are organized according to their frequency tuning ('tonotopically') in the nuclei up to and including the level of the primary auditory field in the forebrain^{21,22}.

The information contained in a single, frequency-specific channel is spatially ambiguous (for example, a given value of ILD at 6 kHz can correspond to sounds from many different locations; Fig. 1c). Therefore, to transform the information about cues into an explicit representation of space, the auditory system integrates information across frequencies and across cues²³, a process that occurs in parallel in the midbrain and forebrain (Fig. 3).

In the midbrain localization pathway, cue information from the tonotopically organized central nucleus of the inferior colliculus (ICC) is combined across frequency channels in the external nucleus of the inferior colliculus (ICX)^{20,24} to create a map of space. This is conveyed to the optic tectum (also called the superior colliculus in mammals), where it contributes to auditory orienting behaviour^{25–27}. In the forebrain localization pathway, cue information is combined across frequency channels beyond the level of the primary auditory field to create clustered representations of space^{20,28,29}. Here, neighbouring neurons are tuned to similar locations or localization cue values, but no global topography exists across a structure. The auditory spatial information in the forebrain pathway contributes to a wide variety of higher-order functions such as working memory, planning complex motor responses and executive control of orienting behaviour^{27,30–32}.

Functional plasticity

Consistent with their effects on auditory orienting behaviour, auditory and visual manipulations cause adaptive changes in the tuning of forebrain and midbrain neurons to sound localization cues. In the forebrain pathway, where space is encoded in a clustered representation, plasticity is usually observed as changes in the distribution of cue values to which neurons are tuned across a large population of

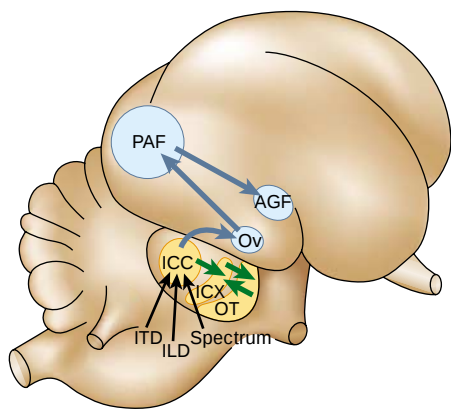


Figure 3 Midbrain and forebrain pathways that mediate auditory orienting responses. This cartoon represents a lateral view of a barn owl's brain. Coloured surfaces represent anatomical structures in the midbrain (yellow) and forebrain (blue). Arrows indicate the flow of information; for clarity, not all connections are shown. ITD, and sound spectrum are processed in parallel in brainstem pathways that project to the central nucleus of the inferior colliculus (ICC) in the midbrain. Spatial information in the ICC is conveyed both to the forebrain, via the thalamic nucleus ovoidalis (Ov), and to the external nucleus of the inferior colliculus (ICX). AGF, archistriatal gaze fields; OT, optic tectum; PAF, primary auditory field.

sampled neurons^{33,34}. In contrast, in the midbrain pathway, because space is encoded as a map, the assay for plasticity is much more precise³⁵. Here, plasticity can be quantified for each neuron as the difference between observed tuning and the tuning predicted by the physical location of the neuron in the nucleus. In the optic tectum, neurons respond to both visual and auditory stimuli and the determination of a neuron's location in the tectum is made accurately from the location of the neuron's visual receptive field, which does not change with experience³⁶. Largely because of the predictive power afforded by mapped representations, studies of the midbrain pathway have provided all that we know about the cellular mechanisms underlying adaptive plasticity in the auditory localization pathway.

Site of plasticity

Before we can explore the mechanisms that underlie behavioural learning, we must determine where in the CNS the cellular changes take place. This has been accomplished in the midbrain localization pathway.

In this pathway, the ICX has been shown to be a site of large-scale adaptive plasticity. In young owls that have experienced either

sustained abnormal hearing or prismatic displacement of the visual field, the tuning of neurons in the ICX and optic tectum to sound localization cues is altered adaptively^{36,37} (Fig. 4). For example, when a young owl experiences a sustained horizontal displacement of the visual field, neurons in the ICX and optic tectum soon begin responding to values of ITD that correspond to a shift in their auditory receptive fields by the amount of the visual field displacement³⁸ (Fig. 4). Gradually, these 'learned responses' become strong, and responses to the normal ITD range, termed 'normal responses', disappear over a period of weeks.

In contrast, the tuning of neurons to the same auditory localization cues remains unchanged in the ICC of prism-reared owls. Thus, in the ICX, where representations of localization cue values are transformed into a map of space, these representations are shaped powerfully by experience.

Associated with the functional changes that take place in the ICX, there is a corresponding change in the anatomy of the axonal projection from the ICC to the ICX^{39,40}. The ICC–ICX projection in normal owls is topographic. In prism-reared owls that have acquired shifted maps of ITD in the ICX, the ICC–ICX projection is broader than normal with bouton-laden axons located both in the normal projection zone and in an abnormal zone where they could support the newly learned ITD tuning (Fig. 5a). The density of the abnormally located axons and boutons exceeds that observed in juvenile owls (Fig. 5b). Therefore, prism experience must induce the formation of learned circuitry in the ICX at least in part through axonal sprouting and synaptogenesis. Interestingly, normal circuitry also persists, showing that alternative learned and normal circuits can coexist in this network.

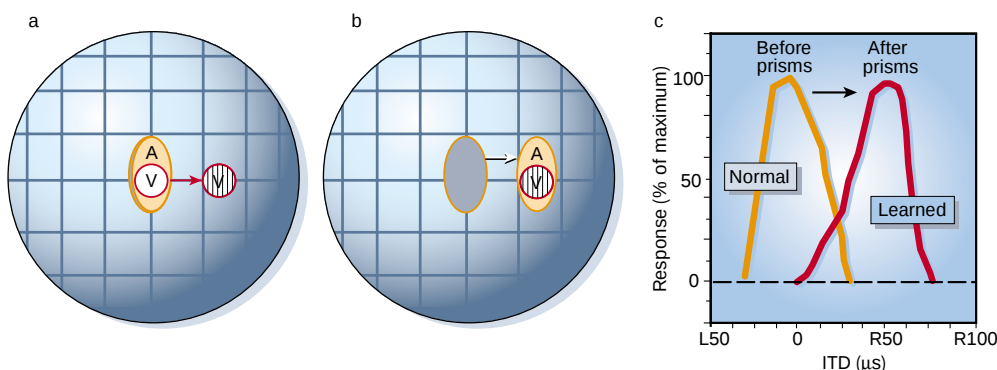
Mechanisms of learning

Given that the ICX is a site where experience-dependent changes take place, what mechanisms, besides the anatomical remodelling discussed above, are involved in the functional plasticity?

NMDA receptors

Neuropharmacological studies have revealed that particular kinds of neurotransmitter receptors contribute critically to the learning process. A special class of glutamate receptor, the *N*-methyl-D-aspartate (NMDA) receptor, is crucial in the expression of newly learned responses. Normal auditory responses in the ICX are driven by glutamatergic synapses, with more than 50% of the synaptic currents provided by NMDA receptors⁴¹ (Fig. 6a). After the first few weeks of prism experience, many ICX neurons express both normal and learned responses. At this stage, when a selective blocker of NMDA receptors (D,L-2-amino-5-phosphonopivalic acid or AP5) is applied focally in the ICX, normal responses decrease by about 50% (as

Figure 4 Plasticity of auditory tuning in the optic tectum of a juvenile barn owl resulting from prism experience. Most neurons in the optic tectum respond to both auditory and visual stimuli and have auditory and visual receptive fields that are mutually aligned in space. **a**, An example of the effect of 23° prisms on the location of a neuron's visual receptive field (encircled V). The globe represents space relative to the owl's line of sight. The auditory receptive field (A) is orange. **b**, After the owl has experienced prisms for 8 weeks, the auditory receptive field has shifted to align with the prismatically displaced visual receptive field. **c**, Plasticity of ITD tuning. ITD tuning of 2 units, both with visual receptive fields located at 0° azimuth with the prisms removed (as shown in **a**), measured before (normal) and after (learned) 8 weeks of prism experience.



L, left ear leading; R, right ear leading.

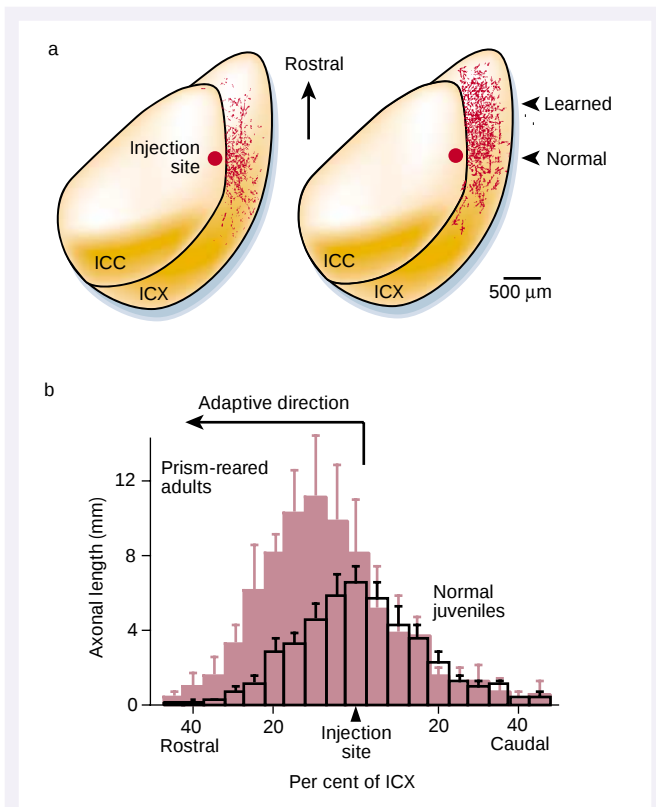


Figure 5 Plasticity of the anatomical projection from the ICC to the ICX, resulting from prism experience. **a**, Digital image drawings of labelled axons in horizontal sections through the ICX after focal injections of a tracer (biocytin) in the ICC. Data from a normal juvenile are shown on the left, whereas the right image shows data from a prism-reared owl with a rostrally shifted map of ITD in the ICX. **b**, Composite spatial distributions of labelled axons for normal juveniles ($n = 7$; open black bars) and prism-reared adults with shifted maps of ITD ($n = 4$; purple bars). Data are from ref. 40.

expected), but learned responses are greatly suppressed and, in some cases, eliminated³⁸. In contrast, when a selective blocker of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors (6-cyano-7-nitroquinoxaline-2,3-dione or CNQX) is applied at the same site, normal responses tend to be suppressed more than learned responses. Thus, in addition to their well-known role in inducing long-term synaptic potentiation in the hippocampus and cerebral cortex, NMDA-receptor currents in the ICX contribute differentially to the functional expression of newly learned responses; this differential contribution disappears over time.

These results indicate that synapses mediating newly learned responses in the ICX have a higher NMDA/AMPA current ratio than synapses mediating normal responses (Fig. 6b). When combined with the anatomical data presented earlier (Fig. 5), these results lead to the intriguing hypothesis that learned responses depend, at least in part, on the formation of new synapses and that these new synapses initially are dominated by NMDA-receptor currents⁴². If true, then as a result of the dependence of NMDA receptors on coincident depolarization and ligand binding for activation⁴³, these synapses would have the advantage that they would transmit information to the postsynaptic neuron only when it was depolarized, perhaps by an instructive signal (see below). Otherwise, the synapses would remain ineffective and so would not disrupt the established pattern of information processing.

GABA_A receptors

Another kind of neurotransmitter receptor, the γ -aminobutyric acid type A (GABA_A) receptor, also contributes importantly to functional

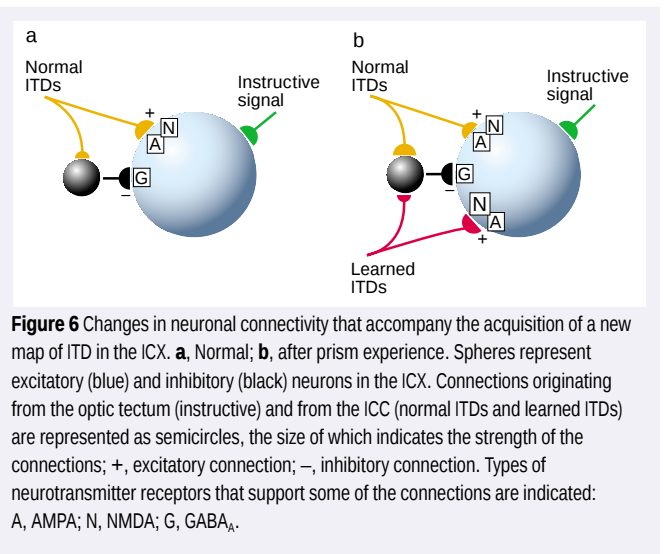


Figure 6 Changes in neuronal connectivity that accompany the acquisition of a new map of ITD in the ICX. **a**, Normal; **b**, after prism experience. Spheres represent excitatory (blue) and inhibitory (black) neurons in the ICX. Connections originating from the optic tectum (instructive) and from the ICC (normal ITDs and learned ITDs) are represented as semicircles, the size of which indicates the strength of the connections; +, excitatory connection; -, inhibitory connection. Types of neurotransmitter receptors that support some of the connections are indicated: A, AMPA; N, NMDA; G, GABA_A.

plasticity. GABA_A-receptor currents inhibit neurons in the ICX⁴⁴. Early in the learning process, strong lateral inhibition mediated by GABA_A receptors suppresses responses to adaptive, newly functional inputs. Blocking GABA_A receptors at this stage causes tuning curves to shift further in the adaptive direction, revealing the full extent of the excitatory plasticity that has occurred⁴⁵. Initially, therefore, GABAergic inhibition masks and, perhaps, opposes excitatory plasticity, thereby preserving the established functional properties of the network.

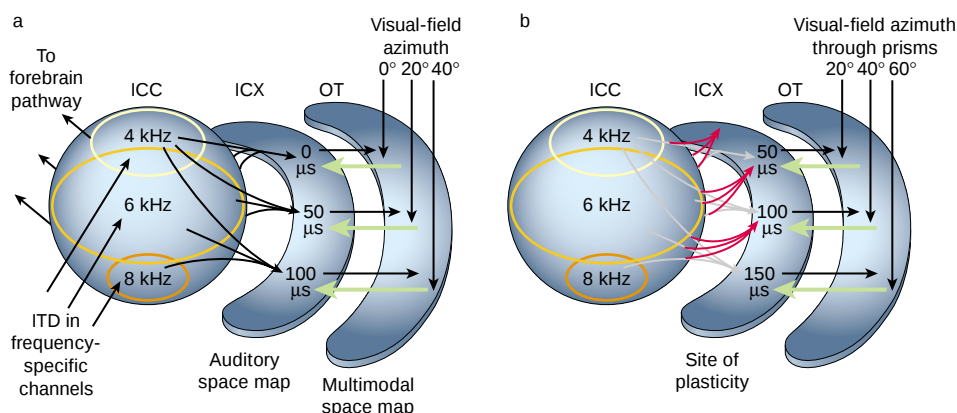
By the end of the learning process, however, GABA_A receptors have a new role⁴⁶. In an ICX that is expressing a fully shifted map of ITD, in which normal responses have been eliminated, focal application of a GABA_A-receptor blocker causes the immediate appearance of normal responses. Thus, in a shifted ITD map, synapses that support normal responses remain active and coexist with synapses that support learned responses (consistent with the anatomy discussed above), but responses to the normal synapses are differentially inhibited by GABAergic inhibition (Fig. 6b; orange connection to black, inhibitory neuron). In this case, GABAergic inhibition eliminates normal responses so that tuning curves shift fully, enabling the selective expression of the learned ITD map. Clearly, changes in the pattern of inhibition, as well as changes in the pattern of excitation, contribute critically to this adaptive plasticity.

The instructive signal

Much of the plasticity that has been observed in the CNS has been induced using paradigms that involve injury to the nervous system, deprivation or excessive use^{47,48}. In these examples, the relative strength of activation across inputs is changed dramatically and the plasticity can be accounted for entirely by competitive, self-organizational forces. But the same is not true for the adaptive adjustment of the auditory space map⁴⁹. In this case, the auditory system is instructed to change its representations of cue values, guided by information provided by a teaching signal. But where does this signal come from and how does it work?

The prism experiments show that the dominant instructive signal in the ICX is provided by the visual system (although other instructive influences exist⁵⁰). The dominance of visual input in this pathway makes sense, in that the primary function of the pathway is to orient gaze towards auditory targets^{26,27}. A visually based signal that calibrates the representation of auditory cues in the ICX could be either a topographic template of the visual field or a foveation-dependent visual signal indicating whether or not auditory orienting responses are accurate⁴⁹. A topographic template signal could instruct changes in the auditory space map by reinforcing auditory synapses that contribute to activity patterns that match those evoked by the visually based template signal and weakening auditory synapses that do not. Alternatively, a

Figure 7 The effect of prism experience on information flow in the midbrain auditory localization pathway. **a**, The pathway in a normal owl; and **b**, in a prism-reared adult with a shifted map of ITD⁴⁰. ITD is measured and mapped in frequency-specific channels in the brainstem. This information ascends to the ICC, and converges across frequency channels in the projection from the ICC to the ICX, where a map of space is created. The map is conveyed to the optic tectum (OT), where it merges with a visual map of space. Green arrows represent the instructive pathway from the OT to the ICX⁵³.



foveation-based instructive signal could guide changes in the auditory space map by strengthening auditory synapses that contribute to orienting movements that cause the stimulus source to fall in the centre of gaze and weakening synapses that do not⁵¹.

A recent experiment has distinguished between these possibilities by exposing owls to different optical conditions at the centre of gaze and in the periphery⁵². A template-based instructive signal predicts different adaptive adjustments in each region of the auditory space map, according to the optical conditions that exist in each region of the visual field. In contrast, a foveation-based instructive signal predicts similar adjustments across the entire space map according to the optical conditions that exist at the centre of gaze. The result of this experiment is that different portions of the auditory space map adjust differently depending on the local visual conditions. Thus, the dominant instructive signal that shapes the auditory space map is a topographic template of visual space.

The source of an instructive signal to the ICX originates from the optic tectum (Figs 6, 7; green connections). Anatomical studies have shown a point-to-point projection from neurons largely in the intermediate layers of the optic tectum back to the ICX⁵³. This projection forms even before owls hatch⁵⁴. The projecting neurons have dendrites that extend into the superficial tectal layers, which receive direct input from the retina, and others that extend into the deep tectal layers, which receive feedforward auditory input from the ICX and visual input from the forebrain. A small lesion placed in the tectum eliminates adaptive plasticity in the corresponding portion of the auditory space map in the ICX, while the rest of the auditory map continues to shift adaptively in response to experience⁵⁵. Similar effects have been observed in ferrets⁵⁶: removal of the superficial layers of the superior colliculus disrupts the development of a normal auditory space map, as assessed in the deep layers of the superior colliculus. The site of plasticity, however, has not been determined in mammals.

The data show that, in barn owls, the optic tectum provides the ICX with a topographic signal that instructs the representation of auditory cue values in the space map (Fig. 7). Because precise, topographic visual activity is strong in the tectal layers where the feedback projection originates³⁵, the tectal signal to the ICX may simply be a retinotopic visual template. But despite numerous attempts, no evidence of a visual instructive signal has been recorded in the ICX of passive, restrained owls, which suggests that the instructive signal might be gated in some fashion, perhaps by attention.

These findings suggest that activity from the visual system is sent into the auditory system as an instructive input to guide the transformation of auditory cue values into a topographic map of space. The same signal is presumably responsible for, and indeed normally used for, adaptive auditory adjustments in response to changes in hearing^{6,57,58}. This visual instructive signal shapes a common repre-

sensation of space for the auditory and visual systems from information that is initially encoded in very different coordinate frames.

A wide variety of transformations in the CNS could be shaped by an analogous strategy. For example, commands for complicated movements are encoded in retinocentric or other sensory frames of reference in some areas of the brain^{59–62}. By studying how the visual system exerts its influence on the auditory space map, we may learn some of the mechanisms by which the nervous system instructs such complex transformations.

Principles of learning

Learning is a balance of innate and experiential influences

The capacity of a network to learn from experience is limited by innate factors that establish and refine initial patterns of connectivity⁶³. Innate patterns of connectivity can contain remarkable specificity, imparting a high degree of functionality that reflects many generations of selection⁶⁴.

The influence of innate patterns of connectivity is apparent in the plasticity of auditory orienting behaviour. For example, owls can alter this behaviour dramatically in response to abnormal sensory conditions only when they are young. In contrast, owls raised with abnormal sensory experience can learn normal orienting behaviour at any age once normal sensory conditions are established¹⁵.

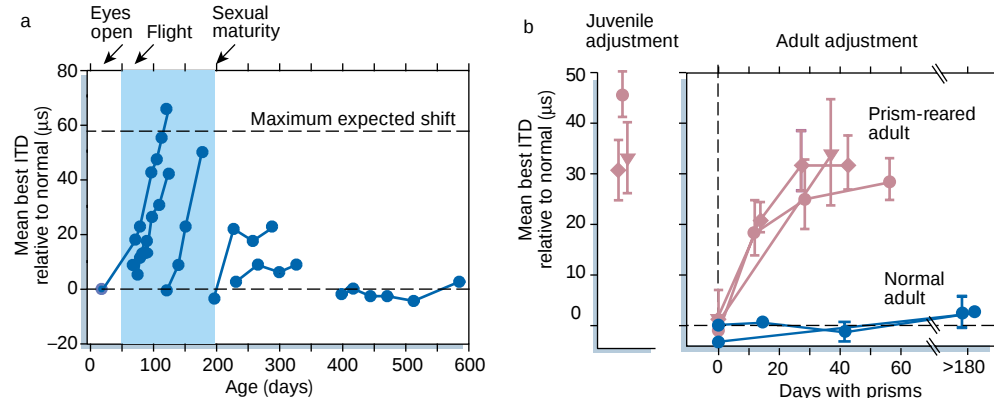
Innate patterns of connectivity are suggested also by the essentially normal maps of ITD that form in owls that are blind from birth or that have been raised with prisms from the day of eye opening^{15,65}. In prism-reared owls, the normal map of ITD that forms initially is subsequently altered by prism experience.

A preference for normal patterns of connections is apparent also in the anatomy and physiology of the ICC–ICX projection. As mentioned previously, in the ICX of prism-reared owls that are expressing shifted maps of ITD, the normal anatomical projection persists along with the learned projection (Fig. 5), even in owls that have never experienced normal correspondences between ITDs and visual locations^{39,40}. Thus, synapses that support normal responses do not require experiential validation in order to become established. Moreover, in shifted maps, the synapses supporting normal responses remain active, but action potentials triggered by them are suppressed by inhibition (Fig. 6b). The converse is not true: in an ICX that has previously expressed a shifted map, blocking inhibition does not cause an immediate re-expression of previously learned responses⁴⁶. Apparently, the strength of learned connections can be reduced to zero through experience, whereas the strength of normal connections cannot. In both respects, synapses that support normal responses seem to be privileged.

Young is better than old

As is true for many behaviours and circuits in the brain, the capacity to change auditory orienting behaviour and the functional

Figure 8 Effect of age and prior experience on ITD map plasticity, as induced by prism experience. **a**, The sensitive period for visual calibration of the ITD map in the optic tectum. These data are from six owls that had 23° prisms mounted over the eyes at different ages. The data points indicate the mean shift of ITD tuning measured for a population of sites. The shaded zone indicates the sensitive period, during which large-scale changes in the ITD map take place under these conditions. Data are from ref. 15. **b**, A trace of juvenile learning persists in adult owls.



Plasticity of the ITD map, induced initially in three juvenile owls (left) by experience with 23° prisms, was re-expressed when these owls were exposed to the same sensory conditions as adults (right, purple symbols). The blue symbols indicate the plasticity of the ITD map in two normally raised owls. All adult owls were over 1 year old when tested. Data are from ref. 67.

properties of neurons in the localization pathway decreases with age^{15,57}. Large-scale adaptive changes in orienting behaviour and in the map of ITD in response to abnormal sensory experience occur in juvenile owls, but not in adult owls under the same conditions (Fig. 8a). A strong age dependence in the plasticity of the auditory space map has also been documented in the superior colliculus of ferrets and guinea pigs^{6,66}.

In barn owls, learning that occurs during the juvenile sensitive period increases the capacity for plasticity in adulthood⁶⁷. An ICX that has acquired an alternative map of ITD during the sensitive period is capable of re-acquiring that map in adulthood, if the same sensory conditions are imposed on the owl (Fig. 8b). The ICX cannot, however, acquire an abnormal map in the adult that has not been learned previously during the juvenile period⁶⁷. Thus, the act of learning an ITD map during the sensitive period leaves a trace in this pathway that endures into adulthood, even though it is not expressed. The plasticity observed in adults reflects the range of learning that occurred during the juvenile period along with innate predispositions. Analogous, long-lasting effects of learning during the sensitive period on adult performance have been reported for song learning in song birds⁶⁸, imprinting in birds and mammals^{69,70}, and language learning in humans⁷¹. Current research is attempting to identify the nature of this persistent learning trace in the owl's auditory localization pathway.

Future directions

Auditory orienting behaviour is a highly quantifiable behaviour that can be shaped powerfully by experience. The pathways in the CNS that contribute to this behaviour are highly conserved across species and are largely identified. In barn owls, the tuning of neurons in these pathways for sound localization cues is unusually sharp and, in the midbrain pathway, the representation of cue values is mapped. Analyses of cellular mechanisms that underlie experience-driven, adaptive changes in orienting behaviour have shown adaptive changes in the nervous system's representation of auditory localization cues, in its anatomical circuitry, and in the contributions of various neurotransmitter-receptor currents to postsynaptic responses. All of these changes are guided by the action of a visually based instructive signal. The results have documented the effects of innate and experiential influences on plasticity, changes in the capacity for plasticity as animals mature, and the enduring effects that early learning can have on adult plasticity.

Future research will delve deeper into the mechanisms of instructed learning in this and in other systems. The cellular and molecular events that underlie learning are known only in rough outline, and

virtually nothing is known about the nature of instructive signals, how they guide plasticity and how they themselves are regulated. The instructive roles played by neuromodulatory systems, such as the cholinergic, noradrenergic and dopaminergic systems that are activated during attention, arousal and learning, are unknown. Also unknown are the factors responsible for the decline in learning capacity with age and the identity of the trace that is left in the adult brain by juvenile learning. From such knowledge, we may discover how to drive learning faster and farther, particularly in adult animals. Finally, research in other species and in other systems will determine the degree to which the mechanisms and principles of learning that have been shown to operate in the auditory localization pathway of barn owls apply generally in the CNS. □

- Field, J., Muir, D., Pilon, R., Sinclair, M. & Dodwell, P. Infants' orientation to lateral sounds from birth to three months. *Child Dev.* **51**, 295–298 (1980).
- Ashmead, D. H., Clifton, R. K. & Reese, E. P. Development of auditory localization in dogs: single source and precedence effect sounds. *Dev. Psychobiol.* **19**, 91–103 (1986).
- Hofman, P. M., Van Riswick, J. G. A. & Van Opstal, A. J. Relearning sound localization with new ears. *Nature Neurosci.* **1**, 417–421 (1998).
- King, A. J., Schnupp, J. W. & Doubell, T. P. The shape of ears to come: dynamic coding of auditory space. *Trends Cogn. Sci.* **5**, 261–270 (2001).
- Knudsen, E. I. Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl. *J. Comp. Physiol.* **185**, 305–321 (1999).
- King, A. J. A map of auditory space in the mammalian brain: neural computation and development. *Exp. Physiol.* **78**, 559–590 (1993).
- Shinn-Cunningham, B. Adapting to remapped auditory localization cues: a decision-theory model. *Percept. Psychophys.* **62**, 33–47 (2000).
- Knudsen, E. I., Blasdel, G. G. & Konishi, M. Sound localization by the barn owl measured with the search coil technique. *J. Comp. Physiol.* **133**, 1–11 (1979).
- Middlebrooks, J. C. & Green, D. M. Sound localization by human listeners. *Annu. Rev. Psychol.* **42**, 135–159 (1991).
- Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E. & Carr, C. E. in *Auditory Function* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 721–745 (Wiley, New York, 1988).
- Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).
- Carr, C. E. & Boudreau, R. E. Development of the time coding pathways in the auditory brainstem of the barn owl. *J. Comp. Neurol.* **373**, 467–483 (1996).
- Knudsen, E. I., Esterly, S. D. & Knudsen, P. F. Monaural occlusion alters sound localization during a sensitive period in the barn owl. *J. Neurosci.* **4**, 1001–1011 (1984).
- Knudsen, E. I., Knudsen, P. F. & Esterly, S. D. A critical period for the recovery of sound localization accuracy following monaural occlusion in the barn owl. *J. Neurosci.* **4**, 1012–1020 (1984).
- Brainard, M. S. & Knudsen, E. I. Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *J. Neurosci.* **18**, 3929–3942 (1998).
- Lackner, J. R. in *Intersensory Perception and Sensory Integration* (eds Walk, R. D. & Pisk, H. L.) 143–173 (Plenum, New York, 1981).
- Knudsen, E. I. & Knudsen, P. F. Visuomotor adaptation to displacing prisms by adult and baby barn owls. *J. Neurosci.* **9**, 3297–3305 (1989).
- Baizer, J. S., Kralj-hans, I. & Glickstein, M. Cerebellar lesions and prism adaptation in macaque monkeys. *J. Neurophysiol.* **81**, 1960–1965 (1999).
- Irvine, D. R. F. in *The Mammalian Auditory Pathway: Neurophysiology* (eds Popper, A. N. & Fay, R. R.) 153–231 (Springer, New York, 1992).
- Cohen, Y. E. & Knudsen, E. I. Maps versus clusters: different representations of auditory space in the midbrain and forebrain. *Trends Neurosci.* **22**, 128–135 (1999).

21. Clarey, J. C., Barone, P. & Imig, T. J. in *The Mammalian Auditory Pathway: Neurophysiology* (eds Popper, A. N. & Fay, R. R.) 232–334 (Springer, New York, 1992).
22. Cohen, Y. E. & Knudsen, E. I. Representation of binaural spatial cues in Field L of the barn owl forebrain. *J. Neurophysiol.* **79**, 879–890 (1998).
23. Brainard, M. D., Knudsen, E. I. & Esterly, S. D. Neural derivation of sound source location: resolution of spatial ambiguity in binaural cues. *J. Acoust. Soc. Am.* **91**, 1015–1027 (1992).
24. Wagner, H., Takahashi, T. & Konishi, M. Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *J. Neurosci.* **7**, 3105–3116 (1987).
25. Stein, B. E. & Meredith, M. A. *The Merging of the Senses* (MIT Press, Cambridge, MA, 1993).
26. Wagner, H. Sound-localization deficits induced by lesions in the barn owl's auditory space map. *J. Neurosci.* **13**, 371–386 (1993).
27. Knudsen, E. I., Knudsen, P. F. & Masino, T. Parallel pathways mediating both sound localization and gaze control in the forebrain and midbrain of the barn owl. *J. Neurosci.* **13**, 2837–2852 (1993).
28. Schreiner, C. E. Order and disorder in auditory cortical maps. *Curr. Opin. Neurobiol.* **5**, 489–496 (1995).
29. Recanzone, G. H. Spatial processing in the auditory cortex of the macaque monkey. *Proc. Natl Acad. Sci. USA* **97**, 11829–11835 (2000).
30. Rauschecker, J. P. & Tian, B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl Acad. Sci. USA* **97**, 11800–11806 (2000).
31. Heffner, H. & Heffner, R. S. Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. *J. Neurophysiol.* **64**, 915–931 (1990).
32. Knudsen, E. I. & Knudsen, P. F. Disruption of auditory spatial working memory by inactivation of the forebrain archistriatum in barn owls. *Nature* **383**, 428–431 (1996).
33. Miller, G. L. & Knudsen, E. I. Early visual experience shapes the representation of auditory space in the forebrain gaze fields of the barn owl. *J. Neurosci.* **19**, 2326–2336 (1999).
34. Miller, G. L. & Knudsen, E. I. Early auditory experience induces frequency-specific, adaptive plasticity in the forebrain gaze fields of the barn owl. *J. Neurophysiol.* **85**, 2184–2194 (2001).
35. Knudsen, E. I. Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.* **2**, 1177–1194 (1982).
36. Brainard, M. S. & Knudsen, E. I. Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *J. Neurosci.* **13**, 4589–4608 (1993).
37. Mogdans, J. & Knudsen, E. I. Early monaural occlusion alters the neural map of interaural level difference in the inferior colliculus of the barn owl. *Brain Res.* **619**, 29–38 (1993).
38. Feldman, D. E. & Knudsen, E. I. Pharmacological specialization of learned auditory responses in the inferior colliculus of the barn owl. *J. Neurosci.* **18**, 3073–3087 (1998).
39. Feldman, D. E. & Knudsen, E. I. An anatomical basis for visual calibration of the auditory space map in the barn owl's midbrain. *J. Neurosci.* **17**, 6820–6837 (1997).
40. DeBello, W. M., Feldman, D. E. & Knudsen, E. I. Adaptive axonal remodeling in the midbrain auditory space map. *J. Neurosci.* **21**, 3161–3174 (2001).
41. Feldman, D. E. & Knudsen, E. I. NMDA and non-NMDA glutamate receptors in auditory transmission in the barn owl inferior colliculus. *J. Neurosci.* **14**, 5939–5958 (1994).
42. Feldman, D. E. & Knudsen, E. I. Experience-dependent plasticity and the maturation of glutamergic synapses. *Neuron* **20**, 1067–1071 (1998).
43. Cotman, C. W., Monaghan, D. T. & Ganong, A. H. Excitatory amino acid neurotransmission: NMDA receptors and Hebb-type synaptic plasticity. *Annu. Rev. Neurosci.* **11**, 61–80 (1988).
44. Fujita, I. & Konishi, M. The role of GABAergic inhibition in processing of interaural time difference in the owl's auditory system. *J. Neurosci.* **11**, 722–729 (1991).
45. Zheng, W. & Knudsen, E. I. GABAergic inhibition antagonizes adaptive adjustment of the owl's auditory space map during the initial phase of plasticity. *J. Neurosci.* **21**, 4356–4365 (2001).
46. Zheng, G. L. & Knudsen, E. I. Functional selection of adaptive auditory space map by GABA-mediated inhibition. *Science* **284**, 962–965 (1999).
47. Kaas, J. H. Plasticity of sensory and motor maps in adult mammals. *Annu. Rev. Neurosci.* **14**, 137–167 (1991).
48. Gilbert, C. D., Sigman, M. & Crist, R. E. The neural basis of perceptual learning. *Neuron* **31**, 681–697 (2001).
49. Knudsen, E. I. Supervised learning in the brain. *J. Neurosci.* **14**, 3985–3997 (1994).
50. Knudsen, E. I. & Mogdans, J. Vision-independent adjustment of unit tuning to sound localization cues in response to monaural occlusion in developing owl tectum. *J. Neurosci.* **12**, 3485–3493 (1992).
51. Rucci, M., Tononi, G. & Edelman, G. M. Registration of neural maps through value-dependent learning: modeling the alignment of auditory and visual maps in the barn owl's optic tectum. *J. Neurosci.* **17**, 334–352 (1997).
52. Hyde, P. S. & Knudsen, E. I. A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum. *J. Neurosci.* **21**, 8586–8593 (2001).
53. Hyde, P. S. & Knudsen, E. I. Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *J. Comp. Neurol.* **421**, 146–160 (2000).
54. Luksch, H., Gauger, B. & Wagner, H. A candidate pathway for a visual instructional signal to the barn owl's auditory system. *J. Neurosci.* **20**, RC70 (2000).
55. Hyde, P. S. & Knudsen, E. I. The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map. *Nature* **415**, 73–76 (2002).
56. King, A. J., Schnupp, J. W. & Thompson, I. D. Signals from the superficial layers of the superior colliculus enable the development of the auditory space map in the deeper layers. *J. Neurosci.* **18**, 9394–9408 (1998).
57. Knudsen, E. I. Experience alters spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl. *J. Neurosci.* **5**, 3094–3109 (1985).
58. Gold, J. I. & Knudsen, E. I. A site of auditory experience-dependent plasticity in the neural representation of auditory space in the barn owl's inferior colliculus. *J. Neurosci.* **20**, 3469–3486 (2000).
59. Batista, A. P., Buneo, C. A., Snyder, L. H. & Anderson, R. A. Reach plans in eye-centered coordinates. *Science* **285**, 257–260 (1999).
60. Cohen, Y. E. & Andersen, R. A. Reaches to sounds encoded in an eye-centered reference frame. *Neuron* **27**, 647–652 (2000).
61. Troyer, T. W. & Doupe, A. J. An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J. Neurophysiol.* **84**, 1204–1223 (2000).
62. Klier, E. M., Wang, H. & Crawford, J. D. The superior colliculus encodes gaze commands in retinal coordinates. *Nature Neurosci.* **4**, 627–632 (2001).
63. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
64. Tinbergen, N. *The Study of Instinct* (Oxford Univ. Press, New York, 1976).
65. Knudsen, E. I., Esterly, S. D. & du Lac, S. Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls: acoustic basis and behavioral correlates. *J. Neurosci.* **11**, 1727–1747 (1991).
66. Withington-Wray, D. J., Binns, K. E. & Keating, M. J. A four-day period of bimodality auditory and visual experience is sufficient to permit normal emergence of the map of auditory space in the guinea pig superior colliculus. *Neurosci. Lett.* **116**, 280–286 (1990).
67. Knudsen, E. I. Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science* **279**, 1531–1533 (1998).
68. Konishi, M. Birdsong: from behavior to neuron. *Annu. Rev. Neurosci.* **8**, 125–170 (1985).
69. Hess, E. H. *Imprinting: Early Experience and the Developmental Psychobiology of Attachment* (Van Nostrand Reinhold, New York, 1973).
70. Scott, J. P. Critical periods in behavioral development. *Science* **138**, 949–958 (1962).
71. Newport, E. L., Bavelier, D. & Neville, H. J. in *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (ed. Doupoux, E.) 481–502 (MIT Press, Cambridge, MA, 2001).
72. Knudsen, E. I. & Knudsen, P. F. Sensitive and critical periods for visual calibration of sound localization by barn owls. *J. Neurosci.* **63**, 131–149 (1990).

Acknowledgements

I thank P. Knudsen for technical support and B. Linkenhoker and Y. Gutfreund for helpful comments on the manuscript. This work was support by grants from the National Institutes of Health, the March of Dimes and the McKnight Foundation.

Circadian rhythms from flies to human

Satchidananda Panda*†, John B. Hogenesch† & Steve A. Kay*†

*Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA (e-mail: stevek@scripps.edu)

†Genomics Institute of Novartis Research Foundation, San Diego, California 92121, USA

In this era of jet travel, our body 'remembers' the previous time zone, such that when we travel, our sleep–wake pattern, mental alertness, eating habits and many other physiological processes temporarily suffer the consequences of time displacement until we adjust to the new time zone. Although the existence of a circadian clock in humans had been postulated for decades, an understanding of the molecular mechanisms has required the full complement of research tools. To gain the initial insights into circadian mechanisms, researchers turned to genetically tractable model organisms such as *Drosophila*.

The rotation of the Earth causes predictable changes in light and temperature in our natural environment. Accordingly, natural selection has favoured the evolution of circadian (from the Latin *circa*, meaning 'about', and *dies*, meaning 'day') clocks or biological clocks — endogenous cellular mechanisms for keeping track of time. These clocks impart a survival advantage by enabling an organism to anticipate daily environmental changes and thus tailor its behaviour and physiology to the appropriate time of the day. The clock is synchronized by the day–night cycle, allowing the organism to accommodate not only the daily cycles of light and dark attributable to the Earth's rotation, but also the alteration in relative span of day and night caused by the tilting of the Earth's axis relative to the Sun. Thus, a circadian timing mechanism that undergoes daily adjustment is useful as a seasonal timer as well.

Constructing a true 24-hour clock, as opposed to a mere sand-timer, is not a trivial task for any organism to undertake during the course of its evolution. Neither is the analysis of how such a precise biological system is assembled and maintained. Genetics of circadian rhythms in flies has elucidated the working principles of the core clock, and provided the tools by which its conservation is also seen in mammals. More recently, genomic analysis of circadian rhythms in flies and mammals has revealed conservation of output physiology that has opened up new avenues in using flies as a model system in the understanding of the daily regulation of behaviour at the molecular level.

Circadian behaviours in *Drosophila*

As several aspects of *Drosophila* physiology and behaviour are restricted to particular times of day, the organism became a natural model system for molecular investigation of circadian regulation. Adult flies emerge from their pupal cases (eclose) when it is cool and moist during the early morning, so minimizing the risk of desiccation as the emerging fly expands its folded wings and hardens its cuticle. Pupae exposed to a 12-h light:12-h dark cycle and subsequently kept in constant darkness also time their eclosion to when they expect dawn (subjective dawn), indicating the



presence of an internal pacemaker¹. Once emerged, adult flies restrict flight, foraging and mating activities to the day (or subjective day), while they tend to 'sleep' (that is, they are relatively unresponsive to sensory stimuli and exhibit rest homeostasis^{2,3}) during the night.

Circadian regulation of such physiology and behaviour results from coordination of the activities of multiple tissues and cell types. An example is the consolidation of feeding behaviour to the day phase, which involves regulation of the sensitivity of chemosensory organs to locate food, activity of the wing muscles to move towards the food, and the action of the digestive system to assimilate nutrients. This locomotor activity rhythm is relatively refractory to acute or minor changes in light levels, such as during lightning and full moons, but is exquisitely sensitive to the timing of dawn and dusk to adapt to the seasonally changing day length.

Discovery of period mutants in *Drosophila*

Early experiments in fly circadian biology established the endogenous nature of the eclosion rhythm. It was found that period length of this rhythm under constant darkness is dependent upon the genetic make-up of the fly strain (reviewed in ref. 1), thus suggesting specific sets of genes were involved in defining complex behaviours (a controversial idea three decades ago). These studies set the stage for a forward genetics approach using eclosion rhythms as a phenotype to identify clock components. Ron Konopka, while a graduate student in Seymour Benzer's laboratory, performed a phenotype-based screen of mutagen-exposed flies and isolated three period mutants in eclosion rhythm, long (*per^L*), short (*per^S*) and arrhythmic (*per⁰*), which mapped to a single genetic locus⁴. These mutant flies exhibited similar defects in locomotor activity rhythms, the detection of which was facilitated by an automated screen developed by Yoshiki Hotta, also in Benzer's laboratory. This more quantitative and persistent trait became the preferred phenotype for mutant screens for the next 30 years (Fig. 1), and a variant of this measure also became a powerful circadian phenotype in rodents.

The pleiotropic effect of each *per* mutation on both eclosion and locomotor activity rhythm established the existence of a single oscillator underlying different rhythms in different developmental stages of the animal. Restoration

of a short-period activity rhythm in *per*⁰ flies by transplantation of adult brain from *per*^S flies on the abdomen of the recipient demonstrated the presence of this oscillator in the fly brain⁵.

Some years later in the mid-1980s, the *period* gene was cloned independently by the Young and Rosbash laboratories^{6,7} and shown to encode a large protein of more than 1,200 amino acids. In transgenic experiments, expression of a wild-type copy of this gene restored normal behavioural rhythms in arrhythmic *per*⁰ flies. Several early studies of these mutants shed light on the role of *per* in maintenance of circadian rhythms. Behavioural analyses of flies harbouring different copy numbers of *per* genes demonstrated dependence of period length on gene dosage—higher doses of *per* decreased period length⁸. Both *per* mRNA and protein levels exhibited rhythmic abundance, which reflected the behavioural rhythm: a near 24-h molecular rhythm in the wild-type flies, no rhythm in *per*⁰ flies, a short period rhythm in *per*^S flies, and a long period rhythm in *per*^L flies.

Subsequent characterization of these flies using genetic mosaic and transgenic approaches defined a group of 20–30 lateral neurons in the adult fly brain as the anatomic site controlling activity rhythm⁹ (Fig. 2). Restricted expression of PER protein in these lateral neurons of *per*⁰ flies restored normal behavioural rhythms¹⁰. Homologues of the *Drosophila per* gene were subsequently cloned from several other insect species^{11–15}, and complementation studies of the *per*⁰ allele with genes derived from other insect species demonstrated conservation of the circadian system and species-specific aspects to its control of rhythmicity¹⁶. Additionally, PER-like immunoreactivity was shown in different orders of animal species^{17,18}, suggesting a functionally conserved clock throughout the animal kingdom. Characterization of *per* in the circadian clock mechanism established the first genetic and anatomic basis for an animal behaviour.

The transcriptional feedback model

The successful identification of *per* spawned subsequent genetic and biochemical screens to identify additional components of the circadian clock. Several additional period mutants were isolated in flies. One of these, *timeless* (*tim*), exhibited characteristics indicative of a true clock component, including all three types of period defect that mapped to the same locus and, more importantly, an elevated level of cytoplasmic PER protein in the arrhythmic allele^{19,20}. Positional cloning was used²¹ to isolate *tim*; it exhibited molecular rhythms similar to PER, and its mRNA and protein levels were coincident with PER in the fly head. TIM protein was also isolated as an interaction partner of PER in a yeast two-hybrid assay²², a method for detecting direct protein–protein interactions.

Detailed molecular genetic characterization of *per* and *tim* offered a skeletal clock mechanism that has been a subject of several

reviews. Both genes are at the core of a transcriptional feedback loop in which their protein products, PER and TIM, dimerize as they accumulate in the cytoplasm during the day, then translocate into the nucleus in the evening to negatively regulate their own transcription. Both proteins are progressively phosphorylated, leading to their eventual degradation in the late night. Neither of these transcriptional inhibitors harbours any domain with similarity to known DNA-binding motifs. However, a domain was identified in PER that shared homology with a second fly protein, the *single-minded* gene product SIM, and a mammalian protein ARNT (aryl hydrocarbon nuclear translocator)^{23,24}. In SIM and ARNT, this PAS (PER, ARNT and SIM) domain is accompanied by a basic region–helix–loop–helix (bHLH) domain, a DNA-binding and heterodimerization surface.

PAS domains were shown to be dimerization surfaces, and given that bHLH proteins often function as heterodimeric pairs (and that PER lacked the bHLH domain), this suggested that PER could function as a transcriptional repressor of a bHLH–PAS heterodimeric pair²⁵. A key element of this mystery was revealed after analyses of *per* and *tim* promoters defined a clock box or E-box (a bHLH protein-binding site) regulatory element that conferred transcriptional cycling^{26,27}. Two additional clock genes were identified in the fly with mammalian orthologues, *dClock* (*dClk*) and *cycle* (*cyc*), which encode proteins containing both bHLH DNA-binding domains and PAS domains^{27–29}. These two proteins were shown to heterodimerize, bind directly to E-box elements found in the *per* and *tim* promoters, and activate their expression. This activation is subsequently inhibited by PER and TIM, thus closing the molecular feedback loop²⁷.

The identification of these new players resulted in the following model for the generation of molecular (and the resultant behavioural) rhythmicity (Fig. 3). The bHLH–PAS heterodimeric pair, dCLK and CYC, reside in the nucleus on the E-box elements in the *per* and *tim* structural genes, positively regulating their transcription. PER and TIM protein levels continue to rise throughout the day to their peak levels in the early evening—a few hours after the peak level of *per* and *tim* mRNAs. The two proteins heterodimerize and translocate into the nucleus where they inhibit the transcriptional activity of the dCLK/CYC complex, thus repressing their own transcription. As both PER and TIM proteins are degraded before dawn, this process is relieved, lifting repression of the dCLK/CYC complex, thereby starting another cycle of PER and TIM accumulation. Incredibly, this core mechanism and several of the above mentioned components are conserved between flies and mammals, over 600 million years of evolutionary time^{27–31}.

Drosophila genetics identified three additional circadian components, *doubletime* (*dbt*), *shaggy* (*sgg*) and *vriille* (*vri*), which act to refine this simple transcriptional–translational feedback loop^{32–34}.

Table 1 Properties of fly and mammalian clock genes and proteins

<i>Drosophila</i> Gene name	Properties	Mammal Gene name	Properties
<i>period</i> (<i>per</i>)	RNA and protein cycle. Binds to TIM. Inhibits dCLK/CYC function.	<i>Period 1</i> (<i>Per1</i>) <i>Period 2</i> (<i>Per2</i>) <i>Period 3</i> (<i>Per3</i>)	RNA and proteins cycle. Physically associates with CRY and among PER proteins. Activator of BMAL1 function. (Mutation in <i>Per2</i> is associated with FASPS.)
<i>timeless</i> (<i>tim</i>)	RNA and protein cycle. Binds to PER and facilitates PER nuclear transport. Inhibits dCLK/CYC function. Degrades in response to light.	<i>Timeless</i> (<i>Tim</i>)	Constitutively expressed. Closest <i>Drosophila</i> relative is <i>timeout</i> . Homozygous null mutant is lethal, making it impossible to conclusively establish clock function.
<i>doubletime</i> (<i>dbt</i>)	Constitutively expressed. Ser/Thr kinase (CK1ε). Phosphorylates TIM-free PER, promoting its degradation.	<i>Casein kinase 1ε</i>	Constitutively expressed. Protein kinase (CK1ε). Phosphorylates PER and affects PER stability.
<i>dClock</i> (<i>dClk</i>)	RNA and protein cycle. bHLH–PAS protein. Heterodimerizes with CYC and promotes transcription from E-box.	<i>Circadian locomotor output cycle kaput</i> (<i>Clock</i>)	Constitutively expressed. Heterodimerizes with BMAL1 and binds to E-box. Promotes transcription of <i>Per</i> and <i>Cry</i> .
<i>cycle</i> (<i>cyc</i>)	Constitutively expressed. bHLH–PAS protein. Heterodimerizes with dCLK and promotes transcription from E-box.	<i>Bmal1/MOP3</i>	RNA cycles. Heterodimerizes with CLOCK and binds to E-box. Promotes transcription of <i>Per</i> and <i>Cry</i> .
<i>cryptochrome</i> (<i>cry</i>)	RNA cycles. Circadian photoreceptor. Promotes light-dependent degradation of TIM. May be essential for some peripheral clocks.	<i>Cryptochrome 1</i> (<i>Cry1</i>) <i>Cryptochrome 2</i> (<i>Cry2</i>)	RNA cycles. Mutations alter rhythmicity in mice, implying a central oscillator function. Physically associates with and stabilizes PER. Inhibits transcription of <i>Per</i> and <i>Cry</i> .
<i>vriille</i>	RNA and protein cycle. bZIP transcription factor. May repress <i>per</i> and <i>tim</i> transcription.	<i>Nfil3/E4BP4</i>	In chicken, represses <i>cPer</i> expression ⁷⁹ . In mouse, suppresses <i>mPer1</i> expression in cell-culture assays ⁷⁹ .
<i>shaggy</i>	Constitutively expressed. Ser/Thr kinase (GSK-3). Promotes TIM phosphorylation and nuclear localization of PER/TIM complex.		

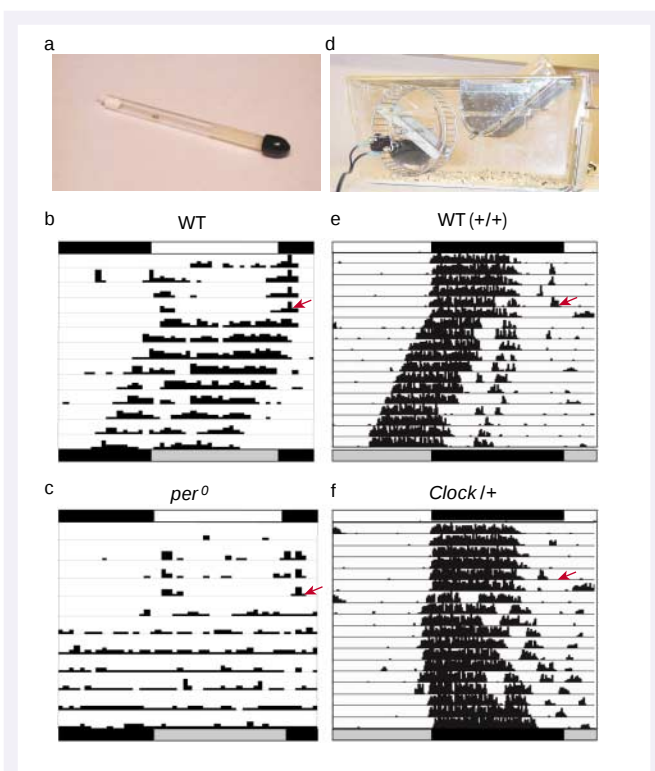


Figure 1 Assay of circadian activity rhythm in flies and mice. **a**, An infra-red beam and optical sensor automatically detect activity of a single fly placed inside a transparent tube and record it on an activity chart or actogram. Actogram of **b**, individual wild-type fly, and **c**, *per*⁰ fly (bearing a loss of function mutation at *per* locus) recorded over several days. The flies were maintained under a 12-h light:12-h dark cycle for few days and then transferred to complete darkness at the time indicated by a red arrow. The *per*⁰ mutation stops the clock and results in arrhythmic activity under constant condition. **d**, Wheel-running activity of individually caged mice gives a measurement of circadian activity rhythm. **e**, Wild-type (+/+) and **f**, mutant mice (+/-) containing one copy of the mutated *Clock* gene both show similar activity pattern under an entraining light:dark condition. After transfer to constant darkness, the wild-type mouse exhibits a rhythm slightly shorter than 24 h, whereas the mutant mouse has a longer period of activity rhythm. Robustness of the assay enabled detection of this mutant in the original screen. Homozygous *Clock/Clock* mice behave in a way that is similar to *per*⁰ flies under extended darkness.

The *dbt* gene encodes the *Drosophila* homologue of the mammalian casein kinase I ϵ (CKI ϵ), and is constitutively expressed³². DBT protein associates physically with both PER and PER/TIM complexes, and may phosphorylate PER³⁵. Outside the PER/TIM complex, phosphorylated PER is unstable. The interplay among PER, TIM and DBT is critical in understanding some human circadian disorders. In early subjective day when monomeric PER is synthesized in the cytoplasm, DBT binds to PER and promotes its phosphorylation, leading to PER degradation and TIM accumulation. High concentrations of TIM promote formation of a stable DBT/PER/TIM complex that can enter the nucleus during early subjective night. Nuclear DBT/PER/TIM complexes are converted to DBT/PER complexes over a period of ~8–10 h. Concomitantly, progressive repression of *per* and *tim* transcription results in decreased accumulation and nuclear entry of the PER/TIM/DBT complex. DBT may progressively phosphorylate PER, leading to its nuclear degradation and contributing to time delays that comprise the circadian transcription–translation feedback loop by either delaying PER accumulation in the cytoplasm or delaying PER's feedback on its own transcription^{35–37}. A second constitutively produced kinase, SGG, promotes TIM phosphorylation, which regulates the timing of nuclear entry of the PER/TIM complex³⁴. Finally, VRI — a basic leucine zipper (bZIP) transcription factor — cycles in the

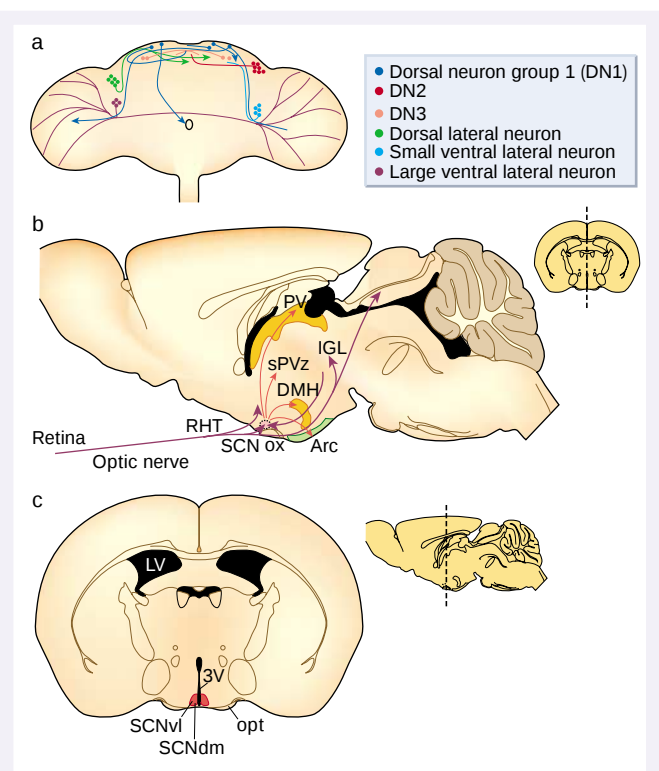


Figure 2 Schematic diagrams showing anatomic features of *Drosophila* and rodent central oscillator. **a**, Diagram of *Drosophila* brain showing the individual neurons expressing the *period* and *timeless* genes, as originally illustrated in ref. 80. Cell bodies are represented as circles, and neurites as lines. The number of circles represents the number of cell bodies, except for DN1 and DN3 (~15 and 40, respectively). Reprinted by permission of Wiley-Liss, Inc. **b**, Longitudinal view of the mouse brain illustrating input pathways to the SCN and outputs from the SCN. Light input from the retina may reach the SCN directly (above the optic chiasm (ox)) via the retinohypothalamic tract (RHT) or indirectly via the intergeniculate leaflet (IGL) of the lateral geniculate nucleus. Brain regions receiving projections from the SCN include subparaventricular zone of the hypothalamus (sPVz), dorsomedial nucleus of the hypothalamus (DMH), paraventricular nucleus of the thalamus (PV), and arcuate hypothalamic nucleus (Arc). These regions in turn mediate many aspects of circadian behaviour and physiology. **c**, Based on differences in morphology, afferent inputs and output projections, the SCN can be divided into the ventrolateral part of the SCN (SCNvl), or 'core' SCN, and the dorsomedial part of the SCN (SCNdm), or 'shell' SCN. See ref. 81 for an exhaustive review of SCN structure and function.

same phase as PER and TIM and has been implicated as a repressor of *per* and *tim* in clock function³³.

Contrary to its name, CYC in the fly does not cycle with any detectable amplitude at the RNA or protein level³⁶. dCLK, however, does cycle with a phase almost opposite to that of PER and TIM³⁹. PER has some role in promoting *dClk* transcription, constituting another feedback loop⁴⁰, although the nature of transcriptional regulation of *dClk* remains a subject of investigation.

Mammalian clockworks

The fundamental anticipatory and light-responsive properties of the circadian pacemaker are conserved between flies and rodents, raising the possibility that the underlying timekeeping mechanism might also be conserved. The first substantial genetic support for this came from a fortuitously isolated rhythm mutant, *tau*, in the golden hamster, which exhibited a short period rhythm in its wheel-running activities⁴¹. The identification of this mutant provided the first genetic tool for the anatomic definition of the circadian pacemaker in a manner similar to that in *Drosophila*. Studies from the early 1970s showed that ablation of the suprachiasmatic nucleus (SCN) — a bilateral pair of hypothalamic nuclei located just above the optic chiasm (Fig. 2) — resulted in

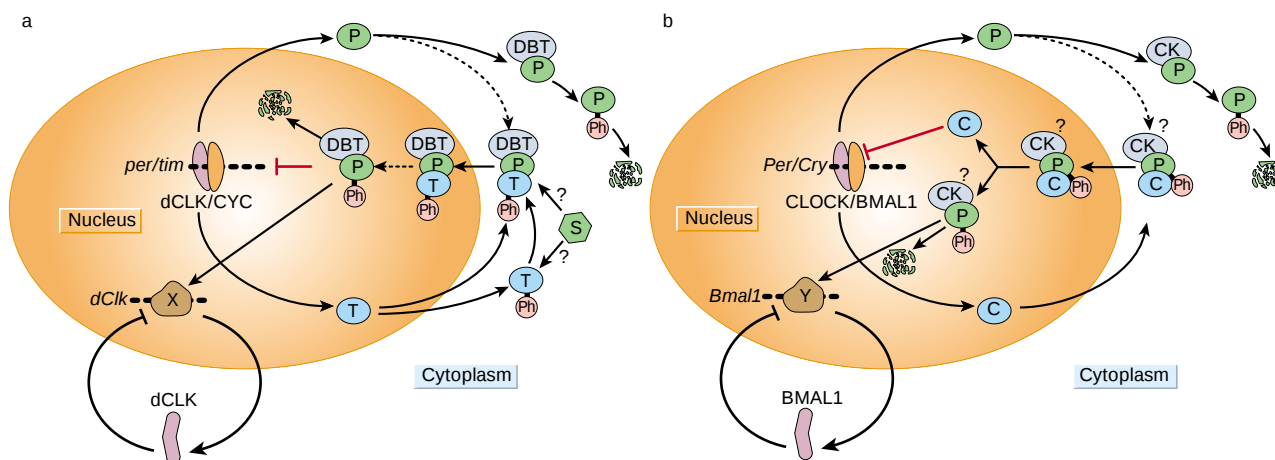


Figure 3 *Drosophila* and mammalian circadian clock. **a**, In *Drosophila*, a heterodimer of two bHLH–PAS domain-containing transcription factors, dCLK and CYC, binds to the E-box in *per* and *tim* promoters, promoting their transcription. DBT phosphorylates (Ph) cytoplasmic TIM-free PER protein (P) and triggers its degradation. As TIM (T) progressively accumulates, it binds to PER, prevents DBT activity, and stabilizes PER. Shaggy (S) phosphorylates TIM, and times the nuclear entry of the PER/TIM/DBT complex. TIM is later released from the nuclear PER/TIM/DBT complex, allowing repression of dCLK/CYC function. In the absence of TIM, DBT promotes the phosphorylation and degradation of nuclear PER, thereby derepressing dCLK/CYC function and starting a new wave of transcription from the E-box. dCLK constitutes another feedback loop by repressing its own transcription. PER promotes *dClk* transcription, although the transcription factor(s) that binds to the *dClk* promoter is currently unknown (X). **b**, In mammals, a heterodimer of two bHLH–PAS domain-containing transcription factors, CLOCK and BMAL1, binds to the E-box in *Per* and *Cry* promoters, and promotes their transcription. CK1 ϵ phosphorylates cytoplasmic PER protein (P) and triggers its degradation. Three different mammalian PER proteins can bind to two mammalian CRY proteins and translocate into the nucleus; here CRY strongly represses CLOCK/BMAL1 activity, whereas PER promotes *Bmal1* transcription. *Bmal1* levels cycle, and the factor(s) that binds to its promoter is currently unknown (Y). Dotted lines represent delays; '?' denotes uncertainty in the step.

complete arrhythmicity of locomotor activity in the rodent⁴². When SCNs from *tau* mutants were transplanted into wild-type hamsters with surgically ablated SCNs, the recipients adopted the short period characteristic of the *tau* mutation. Finally, SCN neurons in culture were shown to have persistent circadian rhythms in their spontaneous firing rate, extending several weeks in culture⁴³.

A combination of forward mutagenesis screening in mice and the use of sequence comparisons with known components of the fly clock has produced a picture of the functional clock in mammals that is highly similar to that in flies (Fig. 3 and Table 1). The most similar components are CLOCK (circadian locomotor output cycle kaput) and BMAL1/MOP3, which are mammalian orthologues of fly dCLK and CYC, respectively. CLOCK and BMAL1/MOP3 were shown to heterodimerize, bind the E-box element (functionally conserved between flies and mammals), and transactivate mammalian genes that harbour this element³¹. The *Clock* mutant (a splice-site mutation resulting in a deletion of a portion of the transactivation surface) reduces *mPer* expression and lengthens the overt activity rhythm (eventually turning arrhythmic)⁴⁴, whereas a loss-of-function *Bmal1/MOP3* mutant abolishes *mPer* expression and eliminates activity rhythms altogether⁴⁵. Mutation of two of the three PER orthologues, *mPer1* and *mPer2*, results in aberrant circadian activity, and the double mutant abolishes rhythmicity^{46,47}.

Although the clock components are conserved across species, their genetic and biochemical roles have diverged. For example, in the mouse *mPER2* seems to activate transcription of *Bmal1*, and exactly opposite to that in flies, BMAL1 cycles in mice whereas CLOCK does not. Therefore, PER positively regulates the rhythmic production of CLOCK/BMAL1 complexes in both mice and flies, although its target has switched. Finally, PER protein products have been shown to weakly suppress CLOCK/BMAL1-dependent *mPer1* transcription in cultured mammalian cells⁴⁴. These results would seem to support a role very similar to that seen for PER (PER/TIM complex) in *Drosophila*, as a negative regulator of its own transcription and a positive regulator of the dCLK/CYC complex. The putative orthologue of *Drosophila timeless*, *mTim*, was found to be a closer orthologue of a second fly gene, *timeout*, which is apparently not

involved in maintenance of circadian rhythmicity. Instead, deletion of the *mTim* gene in the mouse causes lethality⁴⁸. Finally, better repressors of CLOCK/BMAL1 molecular activity were isolated in the orthologues of a *Drosophila* photoreceptor called cryptochrome (CRY; see below).

Resetting the clock

The circadian clock is sensitive to the timing of light exposure. During the middle of subjective day, when light is expected, it has no effect on phase. However, a light pulse administered around subjective dusk (or early night) causes a phase delay, whereas a light pulse near subjective dawn (or early morning) causes a phase advance. This differential sensitivity to light is known as the phase response curve, and is a hallmark of clock function across species. In nature, this property allows the clock to function as a timing device to measure day length, enabling organisms to synchronize their physiology with changing seasons (it also enables jet travellers to adjust to new time zones).

Fly research has provided clues as to how clock resetting occurs in other animals. Stability of TIM protein is light sensitive — even a brief light pulse can trigger its degradation — and this change in TIM level can reset the molecular clock and result in resetting of activity rhythm^{49–51}. Because TIM protein does not possess any chromophore-binding site, the initial step of light perception must be mediated by a photoreceptor. Early experiments showed that rhodopsin is not the circadian photoreceptor in flies, as depletion of the rhodopsin chromophore (vitamin A)⁵² or presence of the *norpA* mutation⁵³ has no effect on entrainment to an external light–dark cycle. Similarly, vitamin A-depleted mice⁵⁴ or mice bearing mutations in visual pathways also exhibit intact circadian entrainment. Even studies in humans have shown that many patients with no significant perception of light as a result of retinal diseases still retain circadian responses to light⁵⁵. Thus, circadian photoperception may have evolved to use a separate mechanism and perhaps separate photoreceptor(s) to filter out weak light stimuli, such as lightning and moonlight, which would otherwise mimic weak light conditions such as dawn and dusk.

A genetic screen for altered rhythmicity identified a *Drosophila* mutant with a light-resetting defect. Designated *cry^{baby}*, the mutant exhibits a normal activity rhythm, entrains properly to a 12-h light:12-h dark cycle, but, unlike wild-type flies, does not phase shift in response to a brief 10-min light pulse during subjective night. Most important, TIM levels are also insensitive to light pulse⁵⁶. Molecular cloning of *cry^{baby}* and subsequent characterization revealed the *dCry* gene encodes a protein that shares extensive sequence similarity with a previously known class of plant circadian photoreceptors, the cryptochromes. The *cry^{baby}* mutation itself corresponds to a highly conserved position in a putative flavin-binding site identified by sequence homology searches⁵⁷.

Biochemical characterization of dCRY showed that it interacted physically with TIM in the yeast two-hybrid system⁵⁸, promoting phosphorylation, ubiquitination and subsequent degradation of TIM by the proteasome⁵⁹. This light-mediated degradation of TIM was shown recently to be dependent on redox activity associated with flavin⁶⁰. Thus, during early subjective night when TIM protein levels rise, light-induced TIM degradation promoted by dCRY delays the accumulation of TIM, which in turn delays the subsequent molecular events of the oscillator machinery, resulting in a phase delay. Conversely, light pulses administered during the late night, when TIM levels are decreasing, facilitates the rapid decline in TIM protein, and causes phase advances. This model assumes no central oscillator role of dCRY, although this protein has recently been implicated in such a role in some peripheral tissues in fly⁶¹.

This clock function of cryptochromes may be conserved in mammals^{62,63}. Although dCRY and its interacting partner TIM are not functionally conserved between flies and mammals, their activity in flies elucidated the integration of two simple molecular mechanisms — a feedback loop and a simple light response — to produce a seemingly complex time-of-day-dependent response of circadian behaviour to light. This also establishes a model for circadian photoresponses in mammals. In contrast to the fly, cryptochrome-deficient mice exhibit circadian rhythm defects, but no conclusive light-resetting defect. Mice deficient in either *mCry1* or *mCry2* exhibit altered period length, and double-mutant mice are completely arrhythmic under constant darkness⁶⁴. Molecular properties of mCRY mirrors fly TIM — cryptochrome mRNA and protein cycle in phase with mPER1 expression in the SCN and retina, mCRY and mPER interact, mCRY proteins inhibit CLOCK/MOP3 transactivation, and double-mutant mice accumulate elevated levels of mPER2, suggesting that the cryptochromes are repressors of mPER expression⁶². In short, cryptochromes have taken over the role of TIM in the mammalian pacemaker.

If CRY took up TIM's job in mice, who is doing CRY's job? The circadian photoreceptor in mammals is yet to be discovered, although the action spectra of circadian resetting support an opsin-based photoreceptor⁶⁵. Genetic analyses have ruled out necessary circadian photoreceptor functions of rod or cone opsins in mammals, although they may have some redundant roles (reviewed in ref. 66).

Clocks at the protein level

For a transcription–translation loop to generate a sustained rhythm, both RNA and protein products of oscillator components must undergo a controlled and rapid degradation. The effect of gene dosage on period length for cycling clock components such as *per* (ref. 8) and constant components such as *Clock* (ref. 67) suggested that tight regulation in their steady-state expression levels is required to generate and sustain a precise rhythm. Early studies showed that the PER and TIM protein products were progressively phosphorylated before degradation. *Drosophila* TIM provided the first clue of how this degradation took place. Pharmacological and *in vitro* studies suggested light-induced tyrosine phosphorylation of TIM is followed by its ubiquitination (a tag for proteasomal degradation)⁵⁹. Identification of the serine/threonine protein kinases DBT and SGG indicated that phosphorylation is essential to trigger degradation of PER and TIM during the circadian cycle. Kinase mutants exhibited

accumulation of the hypophosphorylated form of PER³² or TIM³⁴. The subsequent steps in degradation of PER and TIM are currently unknown. Phosphorylation of other clock components has been reported, but the respective kinases are yet to be discovered^{39,63}.

Transcriptome analysis and behaviour

Although the molecular mechanism by which the central oscillator controls timekeeping is becoming increasingly clear, knowledge of how this timing information is transmitted to regulate behaviour and physiology is only just emerging. A common theme in connecting the clock to physiological outputs has been the identification of cycling component(s), followed by molecular genetic and histological tests to establish a connection. Using this paradigm, two clock-controlled genes, *lark* and *pdf* (whose protein levels, but not RNA, oscillate), were shown to be key mediators of eclosion and activity rhythms in flies^{68,69}. However, how the central oscillator controls rhythmic accumulation of a protein at the post-transcriptional level is entirely unknown.

Modern genomics tools are increasingly important in identification of the transcriptional outputs of the circadian clock. Early success came in a differential display screen for genes expressed in wild-type and clock-deficient flies. This screen identified a gene called *takeout*, whose mRNA is coincidentally expressed with TIM and encodes a lipophilic, ligand-binding protein. The gene is acutely induced in response to starvation in feeding-related organs in insects, and therefore may be important in establishing circadian feeding behaviour⁷⁰. A similar genomics approach identified the *vri* clock component³³.

These successes have encouraged systematic analysis of the circadian pattern of gene expression in different tissue types and genotypes of *Arabidopsis*, flies, mice and rats using DNA microarrays^{71–76}. In each organism, the temporal gene-expression data sets detected cycling of hundreds of transcripts, many times more than previously identified. The list includes already known clock-controlled genes (thus validating the approach), candidate cycling genes involved in known cycling pathways and processes, new processes under circadian control, and key regulators orchestrating coordination of clock-controlled processes. Genomic characterization of mutant flies lacking an essential clock component abolished cycling of all clock-controlled genes, conclusively demonstrating the existence of only one central molecular oscillator in animals^{73,74}.

The scenario seems more complex when we compare the cycling gene sets in different tissue types. Many genes that cycle in fly head do not cycle in the body and vice versa⁷⁵. Similar comparison in more defined tissue types, such as mouse SCN and liver, reveals that most cycling transcripts are tissue specific, implying that circadian transcriptional output functions to temporally regulate physiology to a specific tissue or cell type⁷⁶.

Complex regulation of gene expression

How are different phases of rhythmic gene expression generated from primarily two principal phases of the central clock? A computational approach analysing promoter regions of coordinately regulated transcripts in *Arabidopsis* has identified a *cis*-acting element that specifies the evening phase of cycling⁷¹. Molecular genetic analysis of this element has not only supported its key role in phase determination, but also identified cycling transcription factors binding to it⁷⁷. In flies and mammals, an E-box promoter element has been implicated in rhythmic expression of *per* (and *tim*) phased genes. This E-box element is enriched in the 5'-upstream region of some genes cycling in phase with *per*^{73,74}, indicating that cycling of E-box-containing genes may be controlled directly by the clock components. Both of these studies focus on the regulation of a small subset of circadian output genes, leaving the mechanism of clock-controlled mRNA expression at other phases unknown.

Overall, analysis of the circadian transcriptome is bringing many new challenges to the forefront of research. Which molecular clock outputs participate in transmitting timing information from the

Box 1

Mutant flies and the genetics of human sleep disorders

A popular chronicler⁸² wrote of the *per^S* mutants in flies, "they woke up about five hours too early, they did the same thing for the rest of their lives". The *tau* mutant in hamster behaves similarly — the mutants wake up 2–4 hours early. The mutation was first narrowed down to a chromosomal region in hamster, and the respective mouse chromosome was found to contain a gene similar to fly *dbt*, *CK1 ϵ* , which was later shown to harbour a mutation⁸³. Molecular threading of CK1 ϵ on the crystal structure on the closely related kinase CK1 δ suggests that the *tau* mutation causes a substitution in a highly conserved amino-acid residue involved in specific recognition of an anionic amino acid in target proteins. Similar to its fly counterpart, mammalian CK1 ϵ interacts physically with and phosphorylates mammalian PER both *in vitro* and *in vivo*, and this phosphorylation destabilizes the PER protein. The *tau* mutation diminishes phosphorylation of PER, and therefore enhances PER's stability^{83–85}.

One human sleep disorder, familial advanced sleep phase syndrome (FASPS), is characterized by circadian rhythms resembling those of *per^S* flies and *tau* hamsters. FASPS was linked to a single-gene mutation on human chromosome 2q in one large kindred. Sequence analysis of the candidate gene *hPer2*, resident in this chromosomal region revealed that the mutation⁸⁶, a serine-to-glycine amino-acid substitution at residue 662, was a putative CK1 ϵ recognition site in *hPer2*. This observation almost instantaneously elicited a testable hypothesis. CSNK1 ϵ (RefSeq name of human CK1 ϵ) may phosphorylate hPER2, resulting in its instability. A mutation in *hPer2* leading to reduced phosphorylation by CSNK1 ϵ may result in its stabilization, which in turn may phase advance the clock, perhaps by activating *Bmal1* transcription.

central pacemaker cells in the lateral neurons in flies or SCN in mammals to other brain regions and peripheral organs? How does the circadian clock function to establish the spatiotemporal pattern of gene expression? How are multiple phases of gene expression generated?

Conservation in clock-controlled processes

Annotation of the genes under circadian control reveals the potentially adaptive functions of circadian rhythms that are well conserved through evolution. For example, the circadian oscillator synchronizes the consolidation of feeding behaviour to the activity (wake) phase. Clock regulation of transporters that channel nutrients intracellularly and rate-limiting enzymes of nutrient-utilization pathways occurs in flies and mice, thus coordinating the expression of proteins needed for efficient digestion to the time of day in which feeding occurs. Intermediate products of nutrient metabolism also supply important precursor molecules, which are more fully utilized owing to clock-controlled regulation of enzymes that convert intermediates to their final bioactive form (cholesterol to testosterone, for example). Finally, the act of feeding exposes an organism to various xenobiotics and pathogens. Clock regulation of several intermediate metabolic pathways, which inactivate and promote excretion of several xenobiotics and degradation products of endobiotics, may be an underlying defence mechanism against this chemical stress, initiated secondarily by the feeding behaviour^{72–76}.

A pattern is emerging whereby evolutionary conservation in clock regulation of a specific physiology results from the regulation of key rate-limiting enzymes. For example, the rate-limiting enzymes in the biosynthesis of cholesterol, haem and bioactive amines exhibit circadian rhythms in mRNA accumulation in both flies and mice^{72–74}. The rate-limiting nature of these proteins, coupled with circadian regulation of their transcription, may be an

adaptive mechanism suited for anticipated circadian rhythms in substrate availability or demand for the end product.

The evolutionary conservation in clock regulation of physiology has encouraged researchers to pursue flies as a model system to interrogate the temporal component of learning, alertness and sleep in mammals. Many clock-regulated genes in fly head also have their mammalian counterparts cycling in rodent SCN. The functional significance of such circadian regulation in nervous system and behaviour can be rapidly tested only in a model organism like the fly. For example, transcription of a calcium-activated large potassium channel, *slowpoke*, and its associated protein, slowpoke-binding protein, are clock controlled in flies. The mammalian homologue of *slowpoke*, *mSlo*, also exhibits circadian transcription of its mRNA, and importantly, its expression is enriched in the SCN. The availability of fly mutant stocks and quantitative phenotypic assays established the role of *slowpoke* in regulation of locomotor activity, as *slowpoke*-deficient flies are arrhythmic with no apparent reduction in total activity⁷⁵. If fly–mammal circadian history is our guide, we may yet discover that *mSlo* is an important regulator of activity rhythms in mammals.

From flies to humans

The study of chronobiology in the fly offers an excellent example of how a model organism can facilitate deciphering of the underlying molecular mechanism for a complex trait like the sleep–wake rhythm in humans (Box 1). The rapid progress in this field can justly be attributed to the focus of researchers on the underlying mechanism for generating an overt rhythm (and not the overt rhythm itself). Much of the future of circadian research should focus on connecting the central oscillator to circadian behaviour and physiology. Identification of cycling transcripts has just begun that process, and will be extended by the study of protein and small-molecule rhythms as enabling technologies emerge. The integration of these data will enable a more complete picture of the maintenance of circadian physiology and behaviour. The breadth and depth of circadian regulation now seems to present the perfect example of systems-level biology where a molecular oscillator ticking in a few key neurons in the brain orchestrates a large number of molecules in multiple tissues to generate overt behavioural rhythms.

Mapping the newly identified clock outputs to specific brain regions in flies and connecting them to the master oscillator in lateral neurons will be a step towards understanding how the master oscillator signals to peripheral tissues. Parallel progress in other branches of fly and mammalian neurobiology may help associate a given clock output with a specific clock-controlled behaviour, such as olfaction or feeding. Neuroendocrine signalling is emerging as an important component in the systemic control of clock functions. The systems-level orchestration of circadian physiology is already generating testable hypotheses at a rate that far exceeds current methods to test them in mammals. Model organisms such as the fly offer readily available genetic and genomic tools, rapid generation (or acquisition from public stock centres) of mutants, RNA interference technologies, and automated, quantitative phenotypic assays to rapidly sift through these hypotheses. The exciting possibility that complex behaviours can be described at the molecular level, and are well conserved across species, underscores the importance of the use of model organisms and comparative behavioural genomics. □

- Pittendrigh, C. S. Circadian systems. I. The driving oscillation and its assay in *Drosophila pseudoobscura*. *Proc. Natl Acad. Sci. USA* **58**, 1762–1767 (1967).
- Shaw, P. J., Cirelli, C., Greenspan, R. J. & Tononi, G. Correlates of sleep and waking in *Drosophila melanogaster*. *Science* **287**, 1834–1837 (2000).
- Hendricks, J. C. *et al.* Rest in *Drosophila* is a sleep-like state. *Neuron* **25**, 129–138 (2000).
- Konopka, R. J. & Benzer, S. Clock mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
- Handler, A. M. & Konopka, R. J. Transplantation of a circadian pacemaker in *Drosophila*. *Nature* **279**, 236–238 (1979).
- Bargiello, T. A., Jackson, F. R. & Young, M. W. Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* **312**, 752–754 (1984).
- Reddy, P. *et al.* Molecular analysis of the period locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. *Cell* **38**, 701–710 (1984).

8. Bayliss, M. K., Bargiello, T. A., Jackson, F. R. & Young, M. W. Changes in abundance or structure of the *per* gene product can alter periodicity of the *Drosophila* clock. *Nature* **326**, 390–392 (1987).
9. Ewer, J., Frisch, B., Hamblen-Coyle, M. J., Rosbash, M. & Hall, J. C. Expression of the period clock gene within different cell types in the brain of *Drosophila* adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. *J. Neurosci.* **12**, 3321–3349 (1992).
10. Vossahl, L. B. & Young, M. W. Circadian rhythms in *Drosophila* can be driven by period expression in a restricted group of central brain cells. *Neuron* **15**, 345–360 (1995).
11. Colot, H. V., Hall, J. C. & Rosbash, M. Interspecific comparison of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. *EMBO J.* **7**, 3929–3937 (1988).
12. Reppert, S. M., Tsai, T., Roca, A. L. & Sauman, I. Cloning of a structural and functional homolog of the circadian clock gene period from the giant silkworm *Antheraea pernyi*. *Neuron* **13**, 1167–1176 (1994).
13. Sauman, I. & Reppert, S. M. Circadian clock neurons in the silkworm *Antheraea pernyi*: novel mechanisms of period protein regulation. *Neuron* **17**, 889–900 (1996).
14. Regier, J. C. *et al.* Evolution and phylogenetic utility of the period gene in Lepidoptera. *Mol. Biol. Evol.* **15**, 1172–1182 (1998).
15. Toma, D. P., Bloch, G., Moore, D. & Robinson, G. E. Changes in period mRNA levels in the brain and division of labor in honey bee colonies. *Proc. Natl Acad. Sci. USA* **97**, 6914–6919 (2000).
16. Levine, J. D., Sauman, I., Imbalzano, M., Reppert, S. M. & Jackson, F. R. Period protein from the giant silkworm *Antheraea pernyi* functions as a circadian clock element in *Drosophila melanogaster*. *Neuron* **15**, 147–157 (1995).
17. Siwicki, K. K., Strack, S., Rosbash, M., Hall, J. C. & Jacklet, J. W. An antibody to the *Drosophila* period protein recognizes circadian pacemaker neurons in *Aplysia* and *Bulla*. *Neuron* **3**, 51–58 (1989).
18. Siwicki, K. K., Schwartz, W. J. & Hall, J. C. An antibody to the *Drosophila* period protein labels antigens in the suprachiasmatic nucleus of the rat. *J. Neurogenet.* **8**, 33–42 (1992).
19. Sehgal, A., Price, J. L., Man, B. & Young, M. W. Loss of circadian behavioral rhythms and per RNA oscillations in the *Drosophila* mutant timeless. *Science* **263**, 1603–1606 (1994).
20. Vossahl, L. B., Price, J. L., Sehgal, A., Saez, L. & Young, M. W. Block in nuclear localization of period protein by a second clock mutation, timeless. *Science* **263**, 1606–1609 (1994).
21. Sehgal, A. *et al.* Rhythmic expression of *timeless*: a basis for promoting circadian cycles in *period* gene autoregulation. *Science* **270**, 808–810 (1995).
22. Gekakis, N. *et al.* Isolation of *timeless* by PER protein interaction: defective interaction between *timeless* protein and long-period mutant PER^L. *Science* **270**, 811–815 (1995).
23. Hoffman, E. C. *et al.* Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science* **252**, 954–958 (1991).
24. Crews, S. T., Thomas, J. B. & Goodman, C. S. The *Drosophila* single-minded gene encodes a nuclear protein with sequence similarity to the per gene product. *Cell* **52**, 143–151 (1988).
25. Huang, Z. J., Edery, I. & Rosbash, M. PAS is a dimerization domain common to *Drosophila* Period and several transcription factors. *Nature* **364**, 259–262 (1993).
26. Hao, H., Allen, D. L. & Hardin, P. E. A circadian enhancer mediates PER-dependent mRNA cycling in *Drosophila melanogaster*. *Mol. Cell. Biol.* **17**, 3687–3693 (1997).
27. Darlington, T. K. *et al.* Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim*. *Science* **280**, 1599–1603 (1998).
28. Rutala, J. E. *et al.* CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* *period* and *timeless*. *Cell* **93**, 805–814 (1998).
29. Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M. A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell* **93**, 791–804 (1998).
30. Hogenesch, J. B., Gu, Y. Z., Jain, S. & Bradfield, C. A. The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc. Natl Acad. Sci. USA* **95**, 5474–5479 (1998).
31. Gekakis, N. *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564–1569 (1998).
32. Price, J. L. *et al.* *double-time* is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation. *Cell* **94**, 83–95 (1998).
33. Blau, J. & Young, M. W. Cycling *vrille* expression is required for a functional *Drosophila* clock. *Cell* **99**, 661–671 (1999).
34. Martinek, S., Inonog, S., Manoukian, A. S. & Young, M. W. A role for the segment polarity gene *shaggy*/GSK-3 in the *Drosophila* circadian clock. *Cell* **105**, 769–779 (2001).
35. Kloss, B., Rothenfluh, A., Young, M. W. & Saez, L. Phosphorylation of PERIOD is influenced by cycling physical associations of DOUBLE-TIME, PERIOD, and TIMELESS in the *Drosophila* clock. *Neuron* **30**, 699–706 (2001).
36. Rothenfluh, A., Young, M. W. & Saez, L. A TIMELESS-independent function for PERIOD proteins in the *Drosophila* clock. *Neuron* **26**, 505–514 (2000).
37. Bao, S., Rihel, J., Bjess, E., Fan, J. Y. & Price, J. L. The *Drosophila* *double-time*⁸ mutation delays the nuclear accumulation of *period* protein and affects the feedback regulation of *period* mRNA. *J. Neurosci.* **21**, 7117–7126 (2001).
38. Bae, K., Lee, C., Sidote, D., Chuang, K. Y. & Edery, I. Circadian regulation of a *Drosophila* homolog of the mammalian Clock gene: PER and TIM function as positive regulators. *Mol. Cell. Biol.* **18**, 6142–6151 (1998).
39. Lee, C., Bae, K. & Edery, I. The *Drosophila* CLOCK protein undergoes daily rhythms in abundance, phosphorylation, and interactions with the PER-TIM complex. *Neuron* **21**, 857–867 (1998).
40. Glossop, N. R., Lyons, L. C. & Hardin, P. E. Interlocked feedback loops within the *Drosophila* circadian oscillator. *Science* **286**, 766–768 (1999).
41. Ralph, M. R. & Menaker, M. A mutation of the circadian system in golden hamsters. *Science* **241**, 1225–1227 (1988).
42. Moore, R. Y. & Eichler, V. B. Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain Res.* **42**, 201–206 (1972).
43. Welsh, D. K., Logothetis, D. E., Meister, M. & Reppert, S. M. Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron* **14**, 697–706 (1995).
44. Jin, X. *et al.* A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell* **96**, 57–68 (1999).
45. Bunger, M. K. *et al.* *Mop3* is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009–1017 (2000).
46. Bae, K. *et al.* Differential functions of *mPer1*, *mPer2*, and *mPer3* in the SCN circadian clock. *Neuron* **30**, 525–536 (2001).
47. Zheng, B. *et al.* Nonredundant roles of the *mPer1* and *mPer2* genes in the mammalian circadian clock. *Cell* **105**, 683–694 (2001).
48. Gotter, A. L. *et al.* A time-less function for mouse *Timeless*. *Nature Neurosci.* **3**, 755–756 (2000).
49. Lee, C., Parikh, V., Itskukaichi, T., Bae, K. & Edery, I. Resetting the *Drosophila* clock by photic regulation of PER and a PER-TIM complex. *Science* **271**, 1740–1744 (1996).
50. Myers, M. P., Wager-Smith, K., Rothenfluh-Hilfiker, A. & Young, M. W. Light-induced degradation of TIMELESS and entrainment of the *Drosophila* circadian clock. *Science* **271**, 1736–1740 (1996).
51. Hunter-Ensor, M., Ousley, A. & Sehgal, A. Regulation of the *Drosophila* protein timeless suggests a mechanism for resetting the circadian clock by light. *Cell* **84**, 677–685 (1996).
52. Zimmerman, W. F. & Goldsmith, T. H. Photosensitivity of the circadian rhythm and of visual receptors in carotenoid-depleted *Drosophila*. *Science* **171**, 1167–1169 (1971).
53. Wheeler, D. A., Hamblen-Coyle, M. J., Dushay, M. S. & Hall, J. C. Behavior in light-dark cycles of *Drosophila* mutants that are arrhythmic, blind, or both. *J. Biol. Rhythms* **8**, 67–94 (1993).
54. Thompson, C. L. *et al.* Preservation of light signaling to the suprachiasmatic nucleus in vitamin A-deficient mice. *Proc. Natl Acad. Sci. USA* **98**, 11708–11713 (2001).
55. Czeisler, C. A. *et al.* Suppression of melatonin secretion in some blind patients by exposure to bright light. *N. Engl. J. Med.* **332**, 6–11 (1995).
56. Stanewsky, R. *et al.* The *CRY³* mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell* **95**, 681–692 (1998).
57. Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Rosbash, M. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
58. Ceriani, M. F. *et al.* Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science* **285**, 553–556 (1999).
59. Naidoo, N., Song, W., Hunter-Ensor, M. & Sehgal, A. A role for the proteasome in the light response of the timeless clock protein. *Science* **285**, 1737–1741 (1999).
60. Lin, F. J., Song, W., Meyer-Bernstein, E., Naidoo, N. & Sehgal, A. Photic signaling by cryptochrome in the *Drosophila* circadian system. *Mol. Cell. Biol.* **21**, 7287–7294 (2001).
61. Krishnan, B. *et al.* A new role for cryptochrome in a *Drosophila* circadian oscillator. *Nature* **411**, 313–317 (2001).
62. Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
63. Lee, C., Etchegaray, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
64. van der Horst, G. T. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
65. Takahashi, J. S., DeCoursey, P. J., Bauman, L. & Menaker, M. Spectral sensitivity of a novel photoreceptive system mediating entrainment of mammalian circadian rhythms. *Nature* **308**, 186–188 (1984).
66. Lucas, R. J. *et al.* Identifying the photoreceptive inputs to the mammalian circadian system using transgenic and retinally degenerate mice. *Behav. Brain Res.* **125**, 97–102 (2001).
67. Antoch, M. P. *et al.* Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell* **89**, 655–667 (1997).
68. Newby, L. M. & Jackson, F. R. A new biological rhythm mutant of *Drosophila melanogaster* that identifies a gene with an essential embryonic function. *Genetics* **135**, 1077–1090 (1993).
69. Renn, S. C., Park, J. H., Rosbash, M., Hall, J. C. & Taghert, P. H. A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in *Drosophila*. *Cell* **99**, 791–802 (1999).
70. Sarov-Blat, L., So, W. V., Liu, L. & Rosbash, M. The *Drosophila* *takeout* gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell* **101**, 647–656 (2000).
71. Harmer, S. L. *et al.* Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000).
72. Grundschober, C. *et al.* Circadian regulation of diverse gene products revealed by mRNA expression profiling of synchronized fibroblasts. *J. Biol. Chem.* **276**, 46751–46758 (2001).
73. McDonald, M. J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
74. Claridge-Chang, A. *et al.* Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron* **32**, 657–671 (2001).
75. Ceriani, M. F. *et al.* Genome-wide expression analysis in *Drosophila* predicts genes controlling circadian behavior. (Submitted.)
76. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* (in press).
77. Alabadi, D. *et al.* Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
78. Doi, M., Nakajima, Y., Okano, T. & Fukada, Y. Light-induced phase-delay of the chicken pineal circadian clock is associated with the induction of *E4bp4*, a potential transcriptional repressor of *cPer2* gene. *Proc. Natl Acad. Sci. USA* **98**, 8089–8094 (2001).
79. Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y. & Okamura, H. Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev.* **15**, 995–1006 (2001).
80. Kaneko, M. & Hall, J. C. Neuroanatomy of cells expressing clock genes in *Drosophila*: transgenic manipulation of the period and timeless genes to mark the perikarya of circadian pacemaker neurons and their projections. *J. Comp. Neurol.* **422**, 66–94 (2000).
81. van Esseveldt, K. E., Lehman, M. N. & Boer, G. J. The suprachiasmatic nucleus and the circadian time-keeping system revisited. *Brain Res. Brain Res. Rev.* **33**, 34–77 (2000).
82. Weiner, J. *Time, Love, Memory* (Vintage Books, New York, 1999).
83. Lowrey, P. L. *et al.* Positional synteny cloning and functional characterization of the mammalian circadian mutation *tau*. *Science* **288**, 483–492 (2000).
84. Keesler, G. A. *et al.* Phosphorylation and destabilization of human period 1 clock protein by human casein kinase 1 ϵ . *NeuroReport* **11**, 951–955 (2000).
85. Vielhaber, E., Eide, E., Rivers, A., Gao, Z. H. & Virshup, D. M. Nuclear entry of the circadian regulator mPER1 is controlled by mammalian casein kinase 1 ϵ . *Mol. Cell. Biol.* **20**, 4888–4899 (2000).
86. Toh, K. L. *et al.* An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* **291**, 1040–1043 (2001).

On the scents of smell in the salamander

John S. Kauer

Department of Neuroscience, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA
(e-mail: john.kauer@tufts.edu)

Our sense of smell is based on a remarkable chemical-detection system that possesses high sensitivity, broad discriminability and plastic, yet stable, function. Understanding how olfactory stimuli translate into perception is a problem of daunting complexity. How do odour-coding events in single cells correlate with emergent properties from the ensemble, and with behaviour? For comprehensive descriptions of neural function, analysis must extend from examination of how elemental principles relate to the function of the whole. The tiger salamander has long been used as an experimental model in studies of olfaction, enabling general questions about olfactory function to be approached.

Many experimental preparations provide important information on olfactory function. These include invertebrates such as the fruitfly¹, bee², sphinx moth³, locust⁴, snail⁵, lobster⁶ and crayfish⁷, roundworm⁸ and paramecium⁹, as well as vertebrates that range from fish^{10–13} to amphibians^{14,15} to rodents^{16–22}. An important attribute of any experimental model is the degree to which it offers an opportunity to make direct comparisons not only at different scales of function, but also among different kinds of data — from studies of anatomy, physiology, biochemistry, behaviour and genetics. There are now more than 200 full papers²³ on the olfactory system of salamanders, ~140 of which focus on *Ambystoma tigrinum*, the tiger salamander, and studies cover a period of around 30 years²⁴. These provide a wealth of information using a variety of techniques, from many different functional levels.

A first step in analysing neuronal circuits often involves dissecting, at ever-finer scales, the structural and functional characteristics of the system's elements — the anatomical pathway, single cells, membranes and, ultimately, individual molecules including receptors, channels, transporters and genes. These are the building blocks of complexity and this approach has been particularly successful in numerous systems. Subsequent steps often involve the more difficult task of trying to reassemble the system to understand how the components are integrated to work together. The ultimate challenge for reductionist analysis, however, even after reassembly, lies in characterizing how ensemble properties of the functional whole relate to the specified underlying mechanisms. One analogy for problems confronting the investigator trying to reconcile dissection and reassembly of a neural system is illustrated in Fig. 1. This is taken from an essay by Roger Penrose²⁵ in which he discusses the difficul-

ty in identifying and understanding the locus of the global attribute of 'impossibility' after dissection of the 'impossible triangle'.

'Understanding' a system generally means that meticulous analysis of its elements has led to formulation of a functional description that makes intuitive sense (the visual system encodes colour, which we can see, so finding single cells that respond to colour makes sense) and meshes with well-founded knowledge bases. In a sensory system, 'understanding' allows one to foresee how the modality will encode a new stimulus.

Attainment of understanding for the sense of smell — how the molecular attributes of vapour-phase compounds are encoded, how they are perceived in everyday human experience, and how they generate behaviour — has long been difficult to achieve. For example, the responses in individual neurons are not well correlated with so-called 'odour primaries'²⁶. This lack of an obvious correlation between psychophysical odour experience and neuronal response has focused attention on the global emergent and distributed aspects of olfactory coding. However, this has required the development of technical and cognitive tools, some of which have become available only recently (for a review, see ref. 27).

It is my contention that an understanding of odour coding has begun to emerge only as both analytical and synthetic approaches to function have been applied across and within multiple levels of the system — a goal most rigorously attainable when data from each level are obtained from the same organism. Furthermore, it seems that the relationship between elemental components and ensemble activity in the olfactory pathway is complex, just as analysing the 'impossible triangle' is complex. This is especially true in systems like the olfactory pathway, where representation and manipulation of information appear



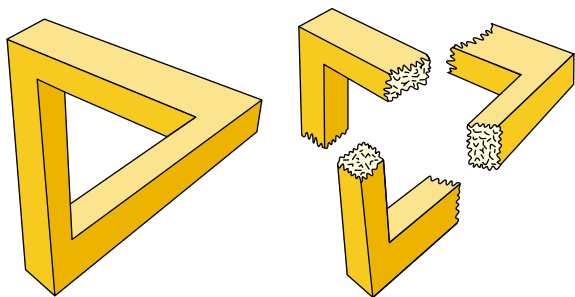


Figure 1 The optical illusion of the impossible triangle illustrates how reductionist dissection to component parts can, in some cases, fail to capture global properties of the whole. (Adapted from ref. 25 with permission.)

to be intricately linked to the disseminated activity that is characteristic of self-organizing structures.

Problems in defining how odours are encoded

Problems in studying olfaction begin with difficulties in defining which aspects of the physical stimuli are encoded by neuronal circuits. Unlike light for the visual system or sound for audition, odours (the perceptual experience) or their constituent compounds, odorants (the physical, vapour-phase molecules), cannot be characterized easily on a continuum of physical and chemical descriptors ordered along well-defined dimensions. For example, molecules that are structurally identical except for their chirality (for example, enantiomers of carvone) can smell different, whereas compounds that are structurally different can smell similar (for example, long chain versus cyclic musks)²⁸. Small changes in molecular structure can generate large changes in perceptual odour experience²⁹.

There are probably thousands or tens of thousands of vapour-phase compounds detectable by olfactory systems, although, interestingly, this number is not known for any species and therefore must be defined for the animal being studied. This means we do not know the scale or grain size of the sensory universe of smell for any animal, including humans. We know that what we can see is within the visual spectrum, but we do not know the extent of what we can smell. From an odour-coding perspective, the precise descriptors of 'odour' space^{30–32} will probably be revealed by how these dimensions are neurally defined by the olfactory pathway and will relate to, but not necessarily be direct reflections of, the dimensionality of 'odorant' space defined by conventional chemical nomenclature. This is interesting because one can imagine developing a system for classifying chemical structure that is based on how odorant descriptors emerge from the olfactory pathway (learning from the biology), just as antigen epitopes can be classified by how they interact with antibodies, rather than by their explicit chemical make-up. All of this emphasizes the importance of correlating any physiological observations, however carefully made, with behaviour from the same animal³³.

These difficulties in identifying underlying dimensions of the olfactory universe are seen in physiological observations. For example, individual odorant sensory cells, and, by inference, the molecular receptors they express^{34,35}, can bind a number of compounds, but apparently with rather moderate affinity^{14,32,36–38}, despite the overall high sensitivity of the system. This reflects the fact that individual receptor molecules seem to be substantially cross-reactive, although probably to varying degrees for different receptors^{39,40}. Indeed, in such a self-organizing, cross-reactive system it is not necessary that the receptor repertoire be defined or even fixed, but only that it encompasses the perceptual odour space required by the animal's behaviour. Many different repertoire sets of different cross-reactive receptors are theoretically capable of encoding all odorants in the same universe³¹. An exception to this

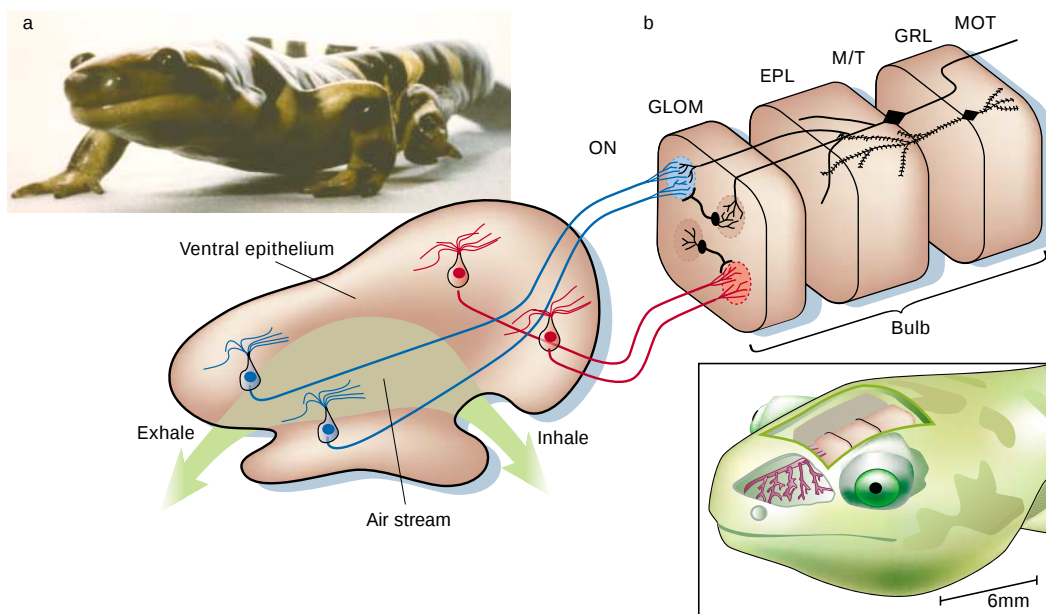


Figure 2 The salamander olfactory pathway. **a**, The tiger salamander *Ambystoma tigrinum*. **b**, A simplified, exploded view of the peripheral components of the salamander olfactory pathway. Schematic in lower-right inset shows the positions of the olfactory epithelium and bulb *in situ*. The larger drawing shows a schematic of the peripheral olfactory circuit, including the bipolar olfactory sensory cells in the epithelium, their axons extending to the bulb (via the olfactory nerve, ON) where they

converge on glomerular targets, and a simplified bulbar circuit. The bulbar circuit comprises the glomerular layer (GLOM), the external plexiform layer (EPL), the mitral/tufted cell layer (M/T) and the granule cell layer (GRL). The medial olfactory tract axons (MOT) project to the higher centres of the central nervous system. (Adapted from ref. 89 with permission.)

may be recognition of pheromones^{41,42}, where high-specificity (and probably high-affinity) binding likely occur.

The difficulties in reconciling odorant structure and neural response first became apparent in observations made 50 years ago by Adrian⁴³. He showed that responses from the olfactory bulb were distributed over space and time, and could discern little correlation between odorant type and response pattern. The introduction of single-cell recording methods allowed examination of relationships between odorants and action potential firing patterns from individual neurons. Gesteland *et al.*⁴⁴ were among the first to apply this technique to the vertebrate olfactory system by recording from olfactory sensory neurons (OSNs). In the frog, they found that single OSNs responded to multiple odorants, could observe no obvious correlations between firing patterns and odorant structure, and concluded that the olfactory system was 'chaos'⁴⁴. Considering their success in correlating stimulus properties with neuronal response in vision⁴⁵, one concern about the olfactory results was whether stimulus delivery had been sufficiently controlled. Would the physiological observations become more 'understandable' (more clearly related to identifiable chemical properties) if odour delivery were better controlled and measured, or did the observed spatial and temporal distributions of response accurately represent the mechanisms by which odours were encoded? All of this emphasized the importance of being able to apply strict spatial, temporal, concentration and quality control over delivered odour stimuli using stimuli that were perceived by the animal.

Why the salamander?

One way to approach these issues for vertebrates was to develop an animal model that possessed the following specific advantages. First, it should be easily obtained and cared for in the laboratory. Second, it must have nasal anatomy that permits precise spatial, temporal, concentration and quality control over odorant presentation, but also have similarities that allow extrapolation to mammals (including receptor-cell turnover^{46,47}). Third, its tissues should have large cells that are tolerant of *in vitro* study, and advantageous optical properties allowing imaging studies; it should permit analyses extending from single-cell biophysics to characterization of ensemble activity distributed across time and space. Finally, the animal must allow both physiological and behavioural responses to odour to be tested and correlated (unlike the frog). After examination of a number of amphibian, reptilian and mammalian candidates, one animal that was found to meet these (and other) requirements was the tiger salamander (Fig. 2a), as originally suggested to me in 1969 by Robert Fink.

The salamander provides advantages for dealing with a number of the challenges in studying smell. For use in the laboratory, salamanders have the practical advantage of being relatively inexpensive and easily obtained from suppliers who collect from the wild. Laboratory husbandry is straightforward. The animal can be bred in captivity⁴⁸ and there is a national facility at the University of Indiana that supports a federally funded programme for supplying, breeding and analysing a related species, the axolotl (*Ambystoma mexicanum*)⁴⁹. A salamander genome project is in its early stages and multitudes of genes are known from retinal and limb libraries⁵⁰.

The anatomy of the tiger salamander nervous system has been studied in an extensive series of classical papers by Herrick⁴⁸ using the Golgi method. Neurons in the brain and olfactory pathway are generally larger than those in mammals and other species, facilitating extracellular, intracellular and patch clamp recording from both primary receptor and bulbar cells^{37,38,51–57} (see generalized circuit in Fig. 2b).

Although still at an early stage of examination, salamander olfactory molecular receptors have been found to be homologous to those in mammals⁵⁸ and *in situ* maps show that the OSNs expressing them are distributed in zonal patterns (J. E. Marchand, A. Strottmann, H. Breer, D. M. Chikaraishi & J.S.K., manuscript in preparation; and

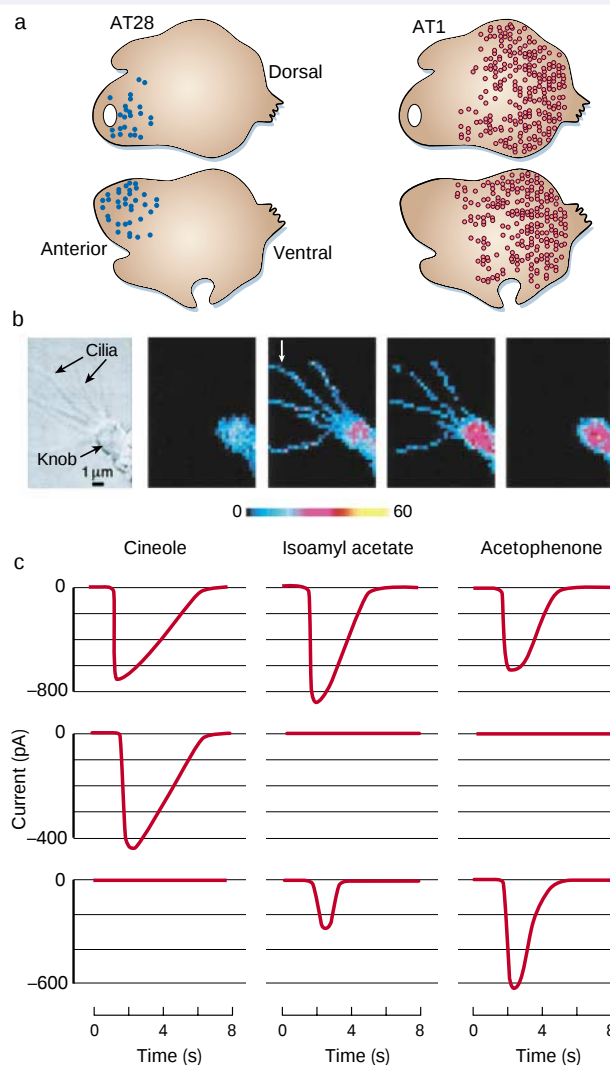


Figure 3 Attributes of salamander olfactory sensory neurons (OSNs). **a**, *In situ* maps of two molecular receptors (AT28 and AT1) show different distributions in the salamander nasal cavity (J. E. Marchand, A. Strottmann, H. Breer, D. M. Chikaraishi & J.S.K., manuscript in preparation). **b**, Imaging of calcium within the cilia and terminal knob of salamander OSNs after odour stimulation shows the progressive changes in intracellular calcium over the course of about 20 s. (Reproduced from ref. 74 with permission.) **c**, Schematic of whole-cell currents from three isolated salamander OSNs showing that different individual cells (rows) can often respond to multiple, unrelated odorants (columns). (Adapted from ref. 37 with permission.)

Figs 3a, 4a). What differs from and is an advantage over mammals is that the nasal cavity is relatively simple, flat and anatomically accessible, permitting correlations of zonal distributions of defined receptor types with neuronal activity (as well as behaviour) recorded from multiple levels of the pathway (Fig. 4). Furthermore, the general connections and circuits of the dorsally located and relatively accessible pyriform cortex (the next relay in the olfactory circuit) have been generally laid out by Herrick⁴⁸, although so far there are no physiological studies of this structure.

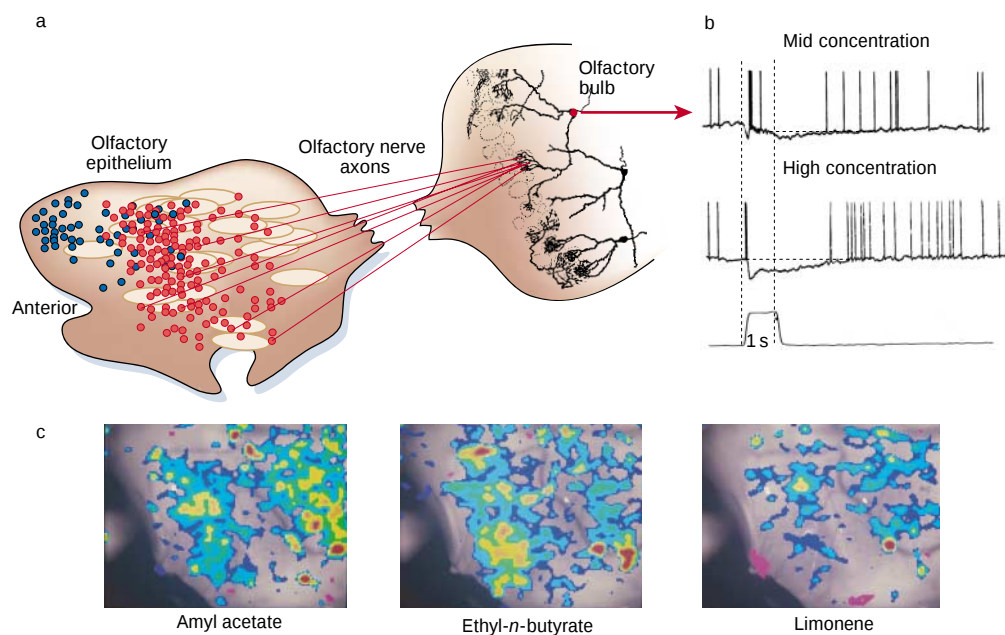
The accessible nature of the nasal cavity allows strict spatial and temporal control of stimuli delivered both to the entire olfactory mucosa and to localized epithelial regions^{37,38,53–55,59}. This degree of spatial control is unlikely ever to be achieved *in situ* in mammals owing to the complex turbinate structure of the nasal cavities. Such experiments have been, and continue to be, crucial for determining relationships among delivered odours, the responding OSN

Figure 4 Correlation of zonal distribution of defined receptor types with receptive fields, single units and distributed signals.

a, Schematic showing the relationships between distributions of two molecular receptors (represented by blue and red coloured dots) expressed in OSNs on the olfactory epithelium (J. E. Marchand, A. Strotman, H. Breer, D. M. Chikaraishi & J.S.K., in preparation) and, from different experiments⁵⁴, the region showing excitatory receptive fields (ovals) for single mitral/tufted (M/T) cells in the salamander olfactory bulb after stimulation with camphor.

b, Intracellular recordings from a single bulbar M/T neuron after 1-s stimulation with amyl acetate at two concentrations. Note the complex changes in temporal firing pattern with a simple

increase in concentration. (Reproduced from ref. 56 with permission.) **c**, The first frame of a real-time sequence of voltage-sensitive dye recordings from one salamander olfactory bulb after 1-s stimulations with three different odorants. Note the different distributed patterns generated by each odorant. (Reproduced from ref. 62 with permission.)



populations, spatial and temporal activity patterns in the circuits of the central nervous system, and output behaviour.

Relatively large signals are obtained using optical recording methods applied to both the olfactory epithelium and bulb, especially with the use of voltage-sensitive dyes that provide simultaneously both high spatial and temporal resolution^{60–64} (Fig. 4c). This degree of temporal resolution using optical recording methods has only recently been achieved in mammalian preparations *in vivo*⁶⁵. In contrast to the radial view of the concentric bulbar layers in mammals, the end-on view of the planar bulbar layers in the *in vivo* salamander (Fig. 2b) permits examination of the temporal evolution of signals both through and across the circuits^{62,63}. Although other optical methods with good spatial resolution have been applied with substantial success to mammalian preparations^{66–68}, the use of voltage-sensitive dyes has permitted analysis of responses in circuit ensembles in relation to the numerous existing electrophysiological studies on receptor and bulbar neurons (Fig. 4b,c). The relatively simple structure of the nasal cavity also has permitted the application of imaging methods to directly observe OSN responses⁶⁴ and to evaluate the effects of flow rate and direction on epithelial odorant response⁶⁹.

Two different behavioural paradigms^{70,71} have shown that the tiger salamander can detect and discriminate odorants (in addition to studies of odour behaviour in certain salamander species in the wild^{72,73}). An immobilized preparation allows odour-related behaviour to be tested by measuring galvanic skin potential⁷¹ while simultaneously recording physiological responses.

Insights into olfactory processing in salamanders

These attributes of the salamander olfactory system have enabled observations about olfactory function to be made at several levels of the pathway. For example, the structural features of OSNs in the olfactory epithelium, as well as their accessibility, relatively large size and position in a simple nasal cavity, have facilitated the generation of *in situ* maps of molecular receptor distributions (J. E. Marchand, A. Strotman, H. Breer, D. M. Chikaraishi & J.S.K., manuscript in preparation; and Figs 3a, 4a), images of changes in calcium concentration within terminal cilia and knobs⁷⁴ (Fig. 3b), and records of

whole-cell currents from isolated cells³⁷ (Fig. 3c). These are only a few of many electrophysiological analyses that have been carried out.

Results from these and other studies^{23,75} have shown how the events involved in olfactory transduction emerge from the following aspects of epithelial physiology: (1) the spatial topography of the nasal cavity where odours interact with different receptor molecules; (2) the initial conductance changes of the receptor membranes; (3) the spread of information, as seen in calcium transients through the cilia and distal knob; and (4) the output spike patterns that are transmitted to the olfactory bulb. The salamander is the only animal in which all of these features have been observed in the same species and can therefore be directly compared with one another.

At the next level, the connections of the olfactory epithelium to the bulb and the integrative events within the bulbar circuits have been the focus of extensive study. There are now numerous recordings of both single-cell activity (Fig. 4b) and ensemble bulbar activity (Fig. 4c), which document a number of important spatial and temporal aspects of responses to odours²³ including how such responses change with concentration and application of odour to localized mucosal sites (receptive-field mapping). Extracellular recordings from the salamander bulb were among the first to show how single units respond to odour stimuli delivered with precise temporal, spatial and concentration control.

Salamanders are the only animals to have been used in the study of receptive-field mapping, and such experiments⁵⁴ were the first to show that distributed OSNs responding to an odorant converge onto single bulbar output neurons. This convergence was further demonstrated in anatomical tracing studies^{76,77}. Recently, molecular marking methods have shown that this unusual anatomical relationship between OSNs and their central targets arises from the fact that distributed OSNs expressing a common molecular receptor send their axons to a few (often one or two) bulbar glomerular targets¹⁸. Although this convergence arrangement was suggested by the earlier studies of receptive fields⁷⁷, its significance for odour coding is still not entirely clear. Although OSNs expressing a single receptor project to defined glomeruli, it is not yet known how this population projects to the different postsynaptic targets within a glomerulus^{61,78}. We also have no information on why projections from the same receptor

population go to multiple, spatially segregated glomeruli. Experiments in which there exists spatially defined access for stimulating specific OSNs *in situ* may be able to address these issues.

As a result of the many different kinds of information available, the salamander provides a unique opportunity for seeing how molecular, anatomical and physiological events relate to one another to form the basis of the odour-coding process that results in output behaviour. One example of such an analysis is shown in Fig. 4a. Here spatial receptive fields of bulbar output neurons for the odorant camphor (ovals) are shown compared to *in situ* maps for two different olfactory receptors (blue and red dots; J. E. Marchand, A. Strottmann, H. Breer, D. M. Chikaraishi & J.S.K., manuscript in preparation). The ovals indicate the epithelial sites at which odour stimulation with camphor could elicit excitatory responses in a bulbar neuron, such as that seen for another odour, amyl acetate (Fig. 4b). In this case, most of the camphor excitatory fields overlaid the distribution of receptor type denoted by red dots, indicating that the receptor mapped by the blue dots contributes little to these receptive-field responses. Although these experiments are only in their initial stages, it is clear that these kinds of comparisons can allow one to generate comprehensive descriptions of relationships between receptor distributions and spatially defined odour responses measured in both single units and in the ensemble circuits. The salamander is the only experimental preparation in which receptor candidates, spatial and temporal control over odour stimulus delivery, single-unit and ensemble recording methods, and odour-guided behaviour are available in one animal.

An overall view of olfactory coding

The kinds of studies described above now form the basis for a large database of directly comparable data of a kind that are difficult or impossible to obtain in other species. Observations from the salamander, together with studies from other preparations, provide strong evidence that the olfactory system is organized to encode odorant molecular structure by systems that generate temporally complex responses that are distributed spatially across multiple cells at each level of the pathway. It is clear that a single odorant type is not encoded by a set of cells with specific responses solely to that compound, with other odorants encoded by activity in different sets of cells. Rather, odorant representation is a combinatorial process involving the use of both neural space and the complex temporal patterning generated by passage of the stimulus through the nasal cavity, linked with integrative mechanisms in the analytical circuits. Data from the multiple levels of analysis possible in the salamander suggested early hypotheses (Fig. 5) about how odour encoding might occur in this animal^{39,61,77}, generating conclusions that are consistent with the continued updating of these hypotheses with data from other species^{16,17,27,40,79,80}.

This view of olfactory function has also been elaborated for other experimental systems, including those in mammals and insects^{4,40}, with different degrees of emphasis placed on spatial versus temporal effects. Dynamical properties of insect and mammalian systems have been examined using both physiological and computational approaches^{4,81–83}. Likewise, details of peripheral olfactory circuits from the salamander have been incorporated into computational analyses^{64,84}. In these studies, spatially and temporally complex responses from simulated bulbar cells have been computed from inferred interactions among the peripheral circuit elements; the computed responses are strikingly similar to those seen *in vivo*.

The salamander is just one of many species contributing to an evolving view of how the component elements of olfactory function are integrated into the odour-coding process. Development of hypotheses about how this process might take place has been greatly facilitated by knowledge gained from other preparations, molecular data on the nature of the receptors and transduction pathways in the primary sensory neurons, detailed information on primary sensory-cell properties, the development of optical methods for examining

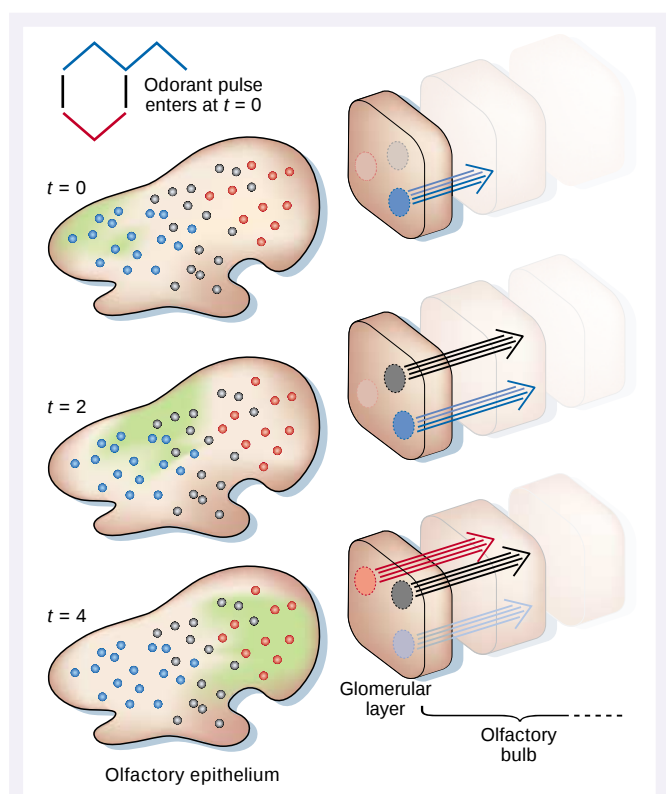


Figure 5 Schematic diagram of spatial and temporal combinatorial coding hypothesis assembled from many different data gathered from the salamander preparation, augmented with information from other preparations (for reviews, see refs 16–18). The simplified olfactory bulb circuit shows the distribution of three populations of OSNs, localized in regions (zones) of the olfactory epithelium, each expressing a single molecular receptor that is sensitive to the same coloured component molecular attributes shown in upper left. Axons (not shown for clarity) from each distributed OSN population converge onto particular glomeruli with associated olfactory bulb circuits. A simulated time series of changes in activation of various OSN and olfactory bulb cell populations is shown as the pulse of odour (shaded region) moves through the nasal cavity. (Adapted from ref. 77 with permission.)

distributed temporal and spatial odour-generated activity in the context of behaviour, and information on the attributes of single-cell responses examined with strict stimulus control. One such scheme is shown in Fig. 5, which in simplified form depicts some of the spatial and temporal encoding events that probably occur in the peripheral olfactory pathway.

Future perspectives

Despite recent advances in characterizing a number of molecular properties of the olfactory system, we have yet to assemble a comprehensive understanding of the steps by which odorants are encoded, or of how odour representation progresses and is modified from one integrative level to the next. For example, we do not know the detailed relationships between the distribution of different molecular receptors in the olfactory epithelium and the odorants that bind to them⁸⁵. Neither do we know the essential functions of bulbar circuits. Data from different species suggest divergent interpretations: is activity sharpened¹⁶ or distributed more broadly⁴? Finally, we have little information about how odours are represented in higher olfactory centres. Although new molecular methods should be helpful for delineating certain anatomical aspects of this part of the pathway⁸⁶, considerable functional analysis remains to be done. The anatomical, physiological and behavioural advantages that the salamander preparation provides should allow investigation of some of these experimental problems.

Much information is still required if we are to understand how global function arises from underlying components and how emergent, self-organizing behaviour comes about. Indeed, there are still problems in reconciling the conundrum of how substantial function is retained after large central nervous system lesions^{87,88}, especially in the context of what we think we know about spatial and temporal patterning. Simple explanations invoking mere redundancy are not adequate if, in the absence of lesions, emphasis is placed on information being critically encoded in spatially defined focal spots of activity. Given the intriguing complexities of how odours are represented by the olfactory pathway, experimental models that permit analysis along the continuum that extends from elemental components to global function should afford advantages for providing insight into how the system works. The salamander preparation has served as one such model.

In addition to contributing to an 'understanding' of the sense of smell, elucidation of olfactory function is important in at least two other contexts. First, these principles may have relevance for understanding other distributed, disseminated brain functions such as memory and learning, which are difficult to access for both technical and conceptual reasons. Second, these biological principles can be used to form the basis for development of biomimetic, artificial devices designed to carry out the same job as the olfactory pathway — that of identifying complex vapour-phase signatures with great sensitivity and high discriminability in noisy environments. Such devices⁶⁴, derived from basic research on appropriate animal models such as the salamander, are beginning to be recognized as important for chemical detection in medical diagnosis, environmental monitoring, and the detection of explosives and other materials related to terrorist activity. □

1. Warr, C., Clyne, P., de Bruyne, M., Kim, J. & Carlson, J. R. Olfaction in *Drosophila*: coding, genetics and e-genetics. *Chem. Senses* **26**, 201–206 (2001).
2. Galizia, C. G., Sachse, S., Rappert, A. & Menzel, R. The glomerular code for odor representation is species specific in the honeybee *Apis mellifera*. *Nature Neurosci.* **2**, 473–478 (1999).
3. Hildebrand, J. G. Analysis of chemical signals by nervous systems. *Proc. Natl Acad. Sci. USA* **92**, 67–74 (1995).
4. Laurent, G. et al. Odor encoding as an active, dynamical process: experiments, computation, and theory. *Annu. Rev. Neurosci.* **24**, 263–297 (2001).
5. Gelperin, A., Tank, D. W. & Tesaro, G. in *Neural Models of Plasticity: Experimental and Theoretical Approaches* (eds Byrne, J. H. & Berry, W. O.) 133–159 (Academic, 1989).
6. Ache, B. W. in *Smell and Taste in Health and Disease* (eds Getchell, T. V., Bartoshuk, L. M., Doty, R. L. & Snow, J. B.) 3–18 (Raven, New York, 1991).
7. Mellon, D. Jr Convergence of multimodal sensory input onto higher-level neurons of the crayfish olfactory pathway. *J. Neurophysiol.* **84**, 3043–3055 (2000).
8. Sengupta, P., Chou, J. H. & Bargmann, C. I. odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84**, 899–909 (1996).
9. Van Houten, J. L., Yang, W. Q. & Bergeron, A. Chemosensory signal transduction in paramecium. *J. Nutr.* **130**, 946S–949S (2000).
10. Ngai, J., Dowling, M. M., Buck, L., Axel, R. & Chess, A. The family of genes encoding odorant receptors in the channel catfish. *Cell* **72**, 657–666 (1993).
11. Restrepo, D., Boekhoff, I. & Breer, H. Rapid kinetic measurements of second messenger formation in olfactory cilia from channel catfish. *Am. J. Physiol.* **264**, C906–C911 (1993).
12. Nikonov, A. A. & Caprio, J. Electrophysiological evidence for a chemotopy of biologically relevant odors in the olfactory bulb of the channel catfish. *J. Neurophysiol.* **86**, 1869–1876 (2001).
13. Friedrich, R. W. & Korsching, S. I. Chemotopic, combinatorial, and noncombinatorial odorant representations in the olfactory bulb revealed using a voltage-sensitive transducer. *J. Neurosci.* **18**, 9977–9988 (1998).
14. Reval, M. F., Sicard, G., Duchamp, A. & Holley, A. New studies on odour discrimination in the frog's olfactory receptor cells. I. Experimental results. *Chem. Senses* **7**, 175–191 (1982).
15. Getchell, T. V. Analysis of unitary spikes recorded extracellularly from frog olfactory receptor cells and axons. *J. Physiol. (Lond.)* **234**, 533–551 (1973).
16. Mori, K., Nagao, H. & Yoshihara, Y. The olfactory bulb: coding and processing of odor molecule information. *Science* **286**, 711–715 (1999).
17. Buck, L. B. Information coding in the vertebrate olfactory system. *Annu. Rev. Neurosci.* **19**, 517–544 (1996).
18. Mombaerts, P. et al. Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
19. Keller, A. et al. Functional organization of rat olfactory bulb glomeruli revealed by optical imaging. *J. Neurosci.* **18**, 2602–2612 (1998).
20. Aroniadou-Anderjaska, V., Ennis, M. & Shipley, M. T. Dendrodendritic recurrent excitation in mitral cells of the rat olfactory bulb. *J. Neurophysiol.* **82**, 489–494 (1999).
21. Wellis, D. P., Scott, J. W. & Harrison, T. A. Discrimination among odorants by single neurons of the rat olfactory bulb. *J. Neurophysiol.* **61**, 1161–1177 (1989).
22. Johnson, B. A. & Leon, M. Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *J. Comp. Neurol.* **422**, 496–509 (2000).
23. Kauer, J. S. Salamander reference list (January 29, 2002) <<http://www.neurosci.tufts.edu>> (2002).
24. Kauer, J. S. *Response Properties of Single Olfactory Bulb Neurons Using Odor Stimulation of Small Nasal Areas in the Salamander*. Thesis, Univ. Pennsylvania (1973).
25. Penrose, R. in *Nature's Imagination: The Frontiers of Scientific Vision* (ed. Cornwell, J.) 12–26 (Oxford Univ. Press, Oxford, 1995).
26. Amore, J. E. The stereochemical theory of olfaction. I. Identification of the seven primary odours. *Proc. Sci. Sect. Toilet Goods Assoc.* **37**, 1–12 (1962).
27. Kauer, J. S. & White, J. Imaging and coding in the olfactory system. *Annu. Rev. Neurosci.* **24**, 963–979 (2001).
28. Beets, M. G. J. The molecular parameters of olfactory response. *Pharm. Rev.* **22**, 1–34 (1970).
29. Polak, E. H. Multiple profile-multiple receptor site model for vertebrate olfaction. *J. Theoret. Biol.* **40**, 469–484 (1973).
30. Alkasab, T. K. et al. Characterizing complex chemosensors: information theoretic analysis of olfactory systems. *Trends Neurosci.* **22**, 102–108 (1999).
31. Alkasab, T. K., White, J. & Kauer, J. S. A computational system for simulating and analyzing arrays of biological and artificial sensors. *Chem. Senses* **27**, 261–275 (2002).
32. Lancet, D., Sadovsky, E. & Seidemann, E. Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system. *Proc. Natl Acad. Sci. USA* **90**, 3715–3719 (1993).
33. Stopfer, M., Bhagavan, S., Smith, B. H. & Laurent, G. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390**, 70–74 (1997).
34. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
35. Touhara, K. et al. Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc. Natl Acad. Sci. USA* **96**, 4040–4045 (1999). [Published erratum in *Proc. Natl Acad. Sci. USA* **97**, 3782 (2000).]
36. Duchamp-Viret, P. & Duchamp, A. Odor processing in the frog olfactory system. *Prog. Neurobiol.* **53**, 561–602 (1997).
37. Firestein, S., Picco, C. & Menini, A. The relation between stimulus and response in olfactory receptor cells of the tiger salamander. *J. Physiol.* **468**, 1–10 (1993).
38. Getchell, T. V. & Shepherd, G. M. Responses of olfactory receptor cells to step pulses of odour at different concentrations in the salamander. *J. Physiol. (Lond.)* **282**, 521–540 (1978).
39. Kauer, J. S. in *Olfaction and Taste* Vol. VII (ed. van der Starre, H.) 227–236 (IRL Press, London, 1980).
40. Hildebrand, J. G. & Shepherd, G. M. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
41. Leinders-Zufall, T. et al. Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792–796 (2000).
42. Kaissling, K. E., Hildebrand, J. G. & Tumlinson, J. H. Pheromone receptor cells in the male moth *Manduca sexta*. *Arch. Insect Biochem. Physiol.* **10**, 273–279 (1989).
43. Adrian, E. D. Sensory messages and sensation. The response of the olfactory organ to different smells. *Acta Physiol. Scand.* **29**, 5–14 (1953).
44. Gesteland, R. C., Lettvin, J. Y. & Pitts, W. H. Chemical transmission in the nose of the frog. *J. Physiol. (Lond.)* **181**, 525–559 (1965).
45. Maturana, H. R., Lettvin, J. Y., McCulloch, W. S. & Pitts, W. H. Anatomy and physiology of vision in the frog. *J. Gen. Physiol.* **43**(Suppl.), 129–175 (1960).
46. Graziadei, P. P. C. & Monti-Graziadei, G. A. in *Handbook of Sensory Physiology* Vol. 9 (ed. Jacobson, M.) 55–83 (Springer, New York, 1978).
47. Simmons, P. A. & Getchell, T. V. Neurogenesis in olfactory epithelium: loss and recovery of transepithelial voltage transients following olfactory nerve section. *J. Neurophysiol.* **45**, 516–528 (1981).
48. Herrick, C. J. *The Brain of the Tiger Salamander* (Univ. Chicago Press, Chicago, 1948).
49. Malacinski, G. M. & Borland, S. J. The Indiana University axolotl colony <<http://www.indiana.edu/~axolotl/>> (2002).
50. Voss, S. R. & Parichy, D. Salamander genome project <<http://lamar.colostate.edu/~svoss/SGP/index.shtml>> (2002).
51. Firestein, S. & Werblin, F. S. Odor-induced membrane currents in vertebrate olfactory receptor neurons. *Science* **244**, 79–82 (1989).
52. Firestein, S., Zufall, F. & Shepherd, G. M. Single odor-sensitive channels in olfactory receptor neurons are also gated by cyclic nucleotides. *J. Neurosci.* **11**, 3565–3572 (1991).
53. Kauer, J. S. Response patterns of amphibian olfactory bulb neurones to odour stimulation. *J. Physiol. (Lond.)* **243**, 695–716 (1974).
54. Kauer, J. S. & Moulton, D. G. Responses of olfactory bulb neurones to odour stimulation of small nasal areas in the salamander. *J. Physiol. (Lond.)* **243**, 717–737 (1974).
55. Kauer, J. S. & Shepherd, G. M. Olfactory stimulation with controlled and monitored step pulses of odor. *Brain Res.* **85**, 108–113 (1975).
56. Hamilton, K. A. & Kauer, J. S. Intracellular potentials of salamander mitral/tufted neurons in response to odor stimulation. *Brain Res.* **338**, 181–185 (1985).
57. Wellis, D. P. & Kauer, J. S. GABA and glutamate receptor involvement in dendrodendritic synaptic interactions from salamander olfactory bulb. *J. Physiol.* **469**, 315–339 (1993).
58. Zhao, H., Dibello, P., Carlson, J. & Firestein, S. PCR amplification of odor receptor cDNA from individual olfactory neurons in salamander. *ACHES Abstr.* (1994).
59. Firestein, S., Shepherd, G. M. & Werblin, F. S. Time course of the membrane current underlying sensory transduction in salamander olfactory receptor neurones. *J. Physiol. (Lond.)* **430**, 135–158 (1990).
60. Kauer, J. S. Real-time imaging of evoked activity in local circuits of the salamander olfactory bulb. *Nature* **331**, 166–168 (1988).
61. Kauer, J. S., Neff, S. R., Hamilton, K. A. & Cinelli, A. R. in *Olfaction as a Model System for Computational Neuroscience* (eds Davis, J. & Eichenbaum, H.) 43–68 (MIT Press, Cambridge, MA, 1991).
62. Cinelli, A. R., Hamilton, K. A. & Kauer, J. S. Salamander olfactory bulb neuronal activity observed by video-rate voltage-sensitive dye imaging. III. Spatio-temporal properties of responses evoked by odorant stimulation. *J. Neurophysiol.* **73**, 2053–2071 (1995).
63. Cinelli, A. R. & Kauer, J. S. Salamander olfactory bulb neuronal activity observed by video-rate voltage-sensitive dye imaging. II. Spatio-temporal properties of responses evoked by electrical stimulation. *J. Neurophysiol.* **73**, 2033–2052 (1995).
64. White, J. & Kauer, J. S. Exploring olfactory population coding using an artificial olfactory system. *Prog. Brain Res.* **130**, 191–203 (2001).

65. Spors, H. & Grinvald, A. Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron* **34**, 301–315 (2002).
66. Uchida, N., Takahashi, Y. K., Tanifuji, M. & Mori, K. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nature Neurosci.* **3**, 1035–1043 (2000).
67. Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
68. Meister, M. & Bonhoeffer, T. Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.* **21**, 1351–1360 (2001).
69. Kent, P. F. & Mozell, M. M. The recording of odorant-induced mucosal activity patterns with a voltage-sensitive dye. *J. Neurophysiol.* **68**, 1804–1819 (1992).
70. Mason, J. R. & Stevens, D. A. Discrimination and generalization among reagent grade odorants by tiger salamanders (*Ambystoma tigrinum*). *Physiol. Behav.* **26**, 647–653 (1981).
71. Dorries, K. M., White, J. & Kauer, J. S. Rapid classical conditioning of odor response in a physiological model for olfactory research, the tiger salamander. *Chem. Senses* **22**, 277–286 (1997).
72. Dawley, E. M. in *Chemical Signals in Vertebrates 4: Ecology, Evolution, and Comparative Biology* (eds Duvall, D., Muller-Schwarze, D. & Silverstein, R. M.) 221–224 (Plenum, New York, 1986).
73. Lindquist, S. B. & Bachmann, M. D. The role of visual and olfactory cues in the prey catching behavior of the tiger salamander, *Ambystoma tigrinum*. *Copeia* **1**, 81–90 (1982).
74. Leinders-Zufall, T., Greer, C. A., Shepherd, G. M. & Zufall, F. Imaging odor-induced calcium transients in single olfactory cilia: specificity of activation and role in transduction. *J. Neurosci.* **18**, 5630–5639 (1998).
75. Lowe, G. & Gold, G. H. Contribution of the ciliary cyclic nucleotide-gated conductance to olfactory transduction in the salamander. *J. Physiol.* **462**, 175–196 (1993).
76. Kauer, J. S. Olfactory receptor cell staining using horseradish peroxidase. *Anat. Rec.* **200**, 331–336 (1981).
77. Kauer, J. S. in *Neurobiology of Taste and Smell* (eds Finger, T. E. & Silver, W. L.) 205–231 (Wiley, New York, 1987).
78. Halasz, N. & Greer, C. A. Terminal arborizations of olfactory nerve fibers in the glomeruli of the olfactory bulb. *J. Comp. Neurol.* **337**, 307–316 (1993).
79. Christensen, T. A. & White, J. in *The Neurobiology of Taste and Smell* (eds Finger, T. E., Silver, W. L. & Restrepo, D.) 197–228 (Wiley-Liss, New York, 2000).
80. Laurent, G. A systems perspective on early olfactory coding. *Science* **286**, 723–728 (1999).
81. Rall, W. & Shepherd, G. M. Theoretical reconstruction of field potentials and dendrodendritic synaptic interactions in the olfactory bulb. *J. Neurophysiol.* **31**, 884–915 (1968).
82. Freeman, W. J. & Baird, B. Relation of olfactory EEG to behavior: spatial analysis. *Behav. Neurosci.* **101**, 393–408 (1987).
83. Freeman, W. J. The physiology of perception. *Sci. Am.* **264**, 78–85 (1991).
84. White, J., Hamilton, K. A., Neff, S. R. & Kauer, J. S. Emergent properties of odor information coding in a representational model of the salamander olfactory bulb. *J. Neurosci.* **12**, 1772–1780 (1992).
85. Scott, J. W., Brierley, T. & Schmidt, F. H. Chemical determinants of the rat electro-olfactogram. *J. Neurosci.* **20**, 4721–4731 (2000).
86. Zou, Z., Horowitz, L. F., Montmayeur, J.-P., Snapper, S. & Buck, L. B. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **414**, 173–179 (2001).
87. Slotnick, B. M., Bell, G. A., Panhuber, H. & Laing, D. G. Detection and discrimination of propionic acid after removal of its 2-DG identified major focus in the olfactory bulb: a psychophysical analysis. *Brain Res.* **762**, 89–96 (1997).
88. Hudson, R. From molecule to mind: the role of experience in shaping olfactory function. *J. Comp. Physiol. A* **185**, 297–304 (1999).
89. Kauer, J. S. Contributions of topography and parallel processing to odor coding in the vertebrate olfactory pathway. *Trends Neurosci.* **14**, 79–85 (1991).

Acknowledgements

Work described in this paper has been generously supported by grants from the NIH (NIDCD), the Office of Naval Research, and the Defense Advanced Research Projects Agency. The studies on the olfactory pathway arising from work in my laboratory have been carried out by many highly skilled, insightful colleagues, including K. Hamilton, A. Cinelli, M. Schwartz Levey, J. White, D. Wellis, K. Dorries, T. Bozza, J. Marchand, T. Alkasab and M.-C. Cheung. I thank B. Talamo and J. White for helpful suggestions on the manuscript.

A small-systems approach to motor pattern generation

Michael P. Nusbaum & Mark P. Beenhakker

Department of Neuroscience, University of Pennsylvania School of Medicine, 215 Stemmler Hall, Philadelphia, Pennsylvania 19104-6074, USA
(e-mail: nusbaum@mail.med.upenn.edu)

How neuronal networks enable animals, humans included, to make coordinated movements is a continuing goal of neuroscience research. The stomatogastric nervous system of decapod crustaceans, which contains a set of distinct but interacting motor circuits, has contributed significantly to the general principles guiding our present understanding of how rhythmic motor circuits operate at the cellular level. This results from a detailed documentation of the circuit dynamics underlying motor pattern generation in this system as well as its modulation by individual transmitters and neurons.

In the quest to understand how the human nervous system enables us to interact with our environment, neuroscientists use a diverse collection of model systems. Investigators working with the mammalian central nervous system (CNS) have made significant progress over the past ten years with respect to the physiological and synaptic properties of individual neurons. However, the large number of neurons and associated complexity of the neuronal networks in these systems have limited progress in understanding how they work. So far, the systems that have provided many of the concepts that guide our understanding of neural circuit operation come from a small number of invertebrate preparations. Among these are neuronal circuits, called central pattern generators (CPGs), that control the generation of rhythmic motor behaviours.

How CPGs generate the patterned neural output underlying rhythmic movements has been studied for nearly 100 years¹. Rhythmic behaviours include all motor acts that at their core involve a rhythmic, repeating set of movements, such as locomotion, respiration and mastication. All CPGs, in both invertebrates and vertebrates, operate on the same general principles²⁻⁴. One such principle is that these networks remain functional in the completely isolated nervous system, in the absence of all rhythmic neuronal input, including the rhythmic feedback from the sensory systems that normally monitor the elicited movements. This property facilitates circuit analysis at the level of the synaptic interactions between the component neurons (that is, cellular-level studies). It should not be forgotten, however, that the ability of sensory and higher-order systems to influence CPG activity is pivotal to ensuring that the resulting behaviour is appropriate for the situation at hand^{2,3,5}. CPGs that are studied in the isolated nervous system include those controlling heartbeat in leeches, feeding in molluscs and crustaceans, locomotion in leeches, molluscs, crustaceans



and various vertebrates, and respiration in mammals²⁻⁴.

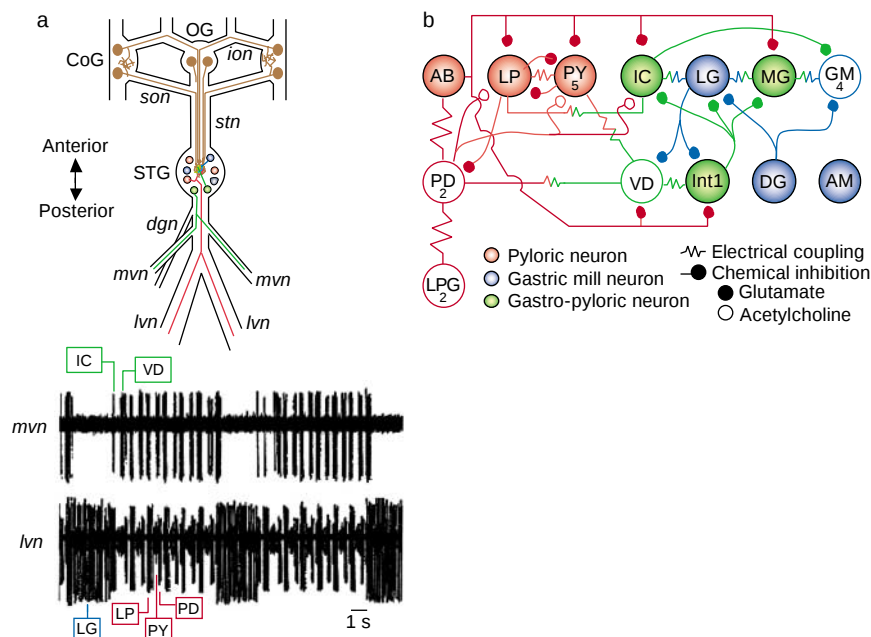
Although the details differ in each circuit, all CPGs use the same set of cellular-level mechanisms for circuit construction²⁻⁴. This includes the prevalence of synaptic inhibition and a set of voltage-dependent firing patterns. Another, more recently appreciated similarity among CPGs is their state dependence. CPG circuits are not dedicated to producing a single neuronal activity pattern. Instead, a CPG is a flexible construct whose output is altered by neuromodulatory transmitters and hormones, enabling them to generate either variants of a single motor pattern (for example, different types of chewing) or, in some cases, distinct activity patterns^{2,3}. This flexibility results largely from the ability of different neuromodulators to change the cellular and synaptic properties of individual circuit neurons. When the properties of circuit components are changed, the output of the circuit itself is modified. These and related aspects of CPG operation have revealed principles that are often shared by non-CPG circuits, enabling CPGs to serve as a model system for understanding neuronal circuit operation in general⁶.

The isolated stomatogastric nervous system

One model system that has contributed importantly to the general understanding of neural circuit operation is the stomatogastric nervous system of decapod crustaceans (lobsters, crabs, crayfish and shrimp). The value of this system has resulted from its accessibility, the use of several innovative techniques and the combined research effort of around 15 laboratories over the past ~30 years^{2,7,8}. Here we highlight the insights gained from studying CPG operation at the cellular level by describing some of the contributions that have come from studying this preparation; a complete list of publications on the stomatogastric system is available at <http://stg.rutgers.edu>.

The stomatogastric nervous system is an extension of the crustacean CNS which contains four ganglia plus their

Figure 1 The gastric mill and pyloric circuits in the stomatogastric nervous system. **a**, Schematic of the stomatogastric nervous system in the crab *Cancer borealis* (top panel). The neurons with somata in the commissural (CoG) and oesophageal (OG) ganglia represent modulatory projection neurons that influence the gastric mill (blue), pyloric (red) and gastro-pyloric (green) neurons in the stomatogastric ganglion (STG). Gastro-pyloric neurons are active with both the gastric mill and the pyloric rhythms. Most STG neurons are motor neurons that project axons posteriorly through the lateral (*lvn*) or medial (*mvn*) ventricular nerves. Illustrated are two STG neurons that project through these nerves. The bottom panel shows that the gastric mill and pyloric rhythms are readily recorded extracellularly via these two nerves *in situ* (recordings modified from ref. 12). Comparable recordings are routinely made *in vitro* (in completely isolated stomatogastric nervous system). The gastric mill rhythm is represented by the relatively long duration, rhythmic action potential bursts in the lateral gastric (LG) neuron and the concomitant elimination of activity in the inferior cardiac (IC) and ventricular dilator (VD) neurons. The faster pyloric rhythm is evident from the sequentially repeating action potential bursts in the lateral pyloric (LP), pyloric (PY), pyloric dilator (PD), IC and VD neurons. The activity in the LP, PY and PD neurons is obscured during each LG neuron burst. Additional abbreviations: *dgn*, dorsal gastric nerve; *ion*, inferior oesophageal nerve; *son*, superior oesophageal nerve; *stn*, stomatogastric nerve. **b**, Schematic of the STG neural network the crab *Cancer borealis*, showing gastric mill and pyloric circuits. The pyloric neurons (red) are shown in their normal sequence of pyloric-timed activity, with time progressing from left to right. The neurons that show gastric mill-timed activity are represented so that the neurons active during protraction of the teeth are on top, and those active during retraction of the teeth are below. Neurons labelled as 'gastro-pyloric neurons' exhibit both gastric mill- and pyloric-timed activity. Abbreviations: AB, anterior burster neuron; MG, medial gastric neuron; GM, gastric mill neuron; Int1, Interneuron 1; DG, dorsal gastric neuron. Adapted from ref. 13.



connecting and peripheral nerves (Fig. 1a). These ganglia include the paired commissural ganglia (~550 neurons each) and the unpaired oesophageal ganglion (12–15 neurons) and stomatogastric ganglion (STG; 25–30 neurons). Within these ganglia are a set of distinct but interacting CPGs that generate the motor rhythms underlying different aspects of feeding in the oesophagus and multi-compartment stomach, including swallowing (oesophagus), food storage (cardiac sac), chewing (gastric mill) and the filtering of chewed food (pylorus)⁷. The two best characterized of these CPGs are the gastric mill and pyloric circuits, both of which are located in the STG. The gastric mill and pyloric motor rhythms have been studied in the intact animal (Fig. 1a), in semi-intact conditions, and in the isolated stomatogastric nervous system^{2,7,9–11}. Although most research in this system has used the isolated nervous system, comparisons with semi-intact and *in vivo* preparations indicate that comparable events occur in all of these conditions^{9,10,12}.

Physiological identification of each gastric mill and pyloric neuron is straightforward because nearly all of the STG neurons are members of one or both of these circuits¹³. All STG neurons are readily recorded and identified in extracellular recordings, enabling a continuous monitor of the activity of all circuit components (Fig. 1a). Additionally, all of these neurons have large-diameter somata (range, 25–120 μm), and each one occurs as either a unique individual or a small group of functionally equivalent neurons. These features enable prolonged (of the order of hours), simultaneous intracellular recordings from several STG neurons, which has greatly facilitated characterization of their synapses and membrane properties, as well as determining the role of each neuron in circuit dynamics. In fact, all synapses among the STG neurons have been identified and characterized, as have the transmitters used by these neurons^{8,14,15} (Fig. 1b). Many of their membrane properties also have been characterized^{15,16}, including endogenous oscillatory capability, plateau potential generation, post-inhibitory rebound and escape from inhibition (Fig. 2).

Two techniques originally developed with the STG system — photoinactivation to selectively delete neurons in a physiological preparation^{14,15,17,18} and the use of the dynamic clamp to selectively reintroduce specific ionic or synaptic currents^{19–21} — were key to obtaining a cellular-level understanding of STG circuit operation. The use of these techniques has also extended well beyond the stomatogastric system^{22–27}.

The gastric mill circuit generates a two-phase rhythm that underlies chewing (protraction and retraction of the teeth) within the gastric mill compartment of the stomach⁹. The rhythmic output of this circuit is an emergent property of the synaptic connectivity and membrane properties of the circuit neurons^{20,28}. The circuit is activated by extrinsic inputs that are usually not spontaneously active^{15,29,30}. In contrast, the pyloric rhythm is 'pacemaker-driven', as the circuit includes an endogenous oscillator neuron (a 'pacemaker neuron') whose intrinsically generated rhythmicity and synaptic outputs enable it to drive this rhythm^{14,17,18}, an ability that is conditional on the presence of one or more neuromodulators. The complete activity pattern generated by the pyloric circuit, however, results from the connectivity and membrane properties of all circuit components¹⁴. As long as the STG continues to receive spontaneously active (non-rhythmic) modulatory input from the commissural ganglia, the pyloric rhythm persists. Pacemaker neurons are also found in other rhythmic systems, including the CPGs underlying mammalian respiration³¹ and locomotion³². This type of neuron seems to be key to the rhythmic activity of the brainstem respiratory circuit^{31,33}, but its role in pattern generation during locomotion remains to be determined.

The conditional membrane and synaptic properties that characterize the STG neurons are regulated by the actions of more than 15 neuromodulatory transmitters contained in the STG terminals of the sensory and projection neurons that innervate this ganglion^{8,34,35} (Fig. 3a). There are several documented cases of co-localized

transmitters within these neurons^{8,11,34} (Fig. 3b), and many of these same modulatory substances, plus some additional ones, also influence the STG as circulating hormones^{8,35}. Application of particular modulators often accelerates and strengthens the pyloric rhythm, with each one eliciting a different version^{2,35} (Fig. 3a), although a few modulators instead inhibit the rhythmic activity^{8,15}. Different modulatory projection neurons also elicit different versions of the pyloric and gastric mill rhythms^{11,30,36,37} (Fig. 3b).

Neural output from the STG circuits is not always followed faithfully by the target muscles. The striated muscles of the stomatogastric system are 'slow' non-twitch muscles that generally respond in a graded fashion to their excitatory input. Although some muscle types do faithfully replicate the pattern of their neural input, others have a sufficiently slow contractile response that they exhibit a tonic contraction despite receiving a rhythmic neural input³⁸. Furthermore, although single motor neurons innervate several different muscles in this system, the individual muscles can generate distinct contraction patterns despite their common innervation³⁸. An active reinterpretation of the neural input at the muscle level has been noted in other systems³⁹, and there is also plasticity at the periphery, increasing even further the potential outputs generated by these motor systems^{40–42}.

Insight into neuronal circuit operation has also come from studying the stomatogastric system in several different crustacean species^{43,44}. One conclusion drawn from these studies is that the functionally equivalent neuron in different species can differ in transmitter phenotype^{8,45} and/or physiological actions^{15,45,46}. This work has implications for work in the vertebrate CNS where the same neurons are studied in different species without the availability of an extended cellular characterization in each case.

Basic circuit operation

Synaptic transmission and membrane properties

In the 1960s and early 1970s, it was believed that the neuronal basis of behaviour could be explained once all of the neurons in a circuit were identified and their synaptic connectivity determined. However, as CPG circuit analyses progressed, it became evident that there was more to motor pattern generation than the presence of excitation and inhibition. Specifically, neuronal membrane properties were found to be equally important for circuit operation^{14,16}. Furthermore, we learned that neuromodulation changed these membrane properties and synaptic strengths, enabling neural circuits to generate multiple activity patterns^{14,15,47}. It is now clear that these influences are key to the operation of all neural circuits.

One feature shared by many CPG circuits is the prevalence of inhibitory synapses^{2,3}. Within the STG circuits, all transmitter-mediated actions are inhibitory¹⁶ (Fig. 1b). During the gastric mill and pyloric rhythms, these inhibitory synapses are responsible for the inactive phase of most STG neurons, whereas their active phase results from the activation of voltage-dependent membrane properties that are primed by the preceding inhibition^{2,14,16}. For example, the rate of post-inhibitory rebound in different pyloric neurons contributes importantly to the timing of their activity within the pyloric rhythm⁴⁸, whereas plateau potential generation strengthens the intensity of the action potential bursts in many pyloric and gastric mill neurons^{15,16}. The interplay between synapses and membrane properties is further exemplified by the fact that the rhythmic synaptic inhibition in the pyloric circuit restricts the burst duration of at least some pyloric neurons to their period of most intense activity⁴⁹.

Another functional consequence of the particular cellular properties associated with STG circuit neurons is that the pyloric rhythm retains its essential features, such as the relative timing of action potential bursts (activity phase within the rhythm) in each neuron, across a broad range of frequencies (~0.2–2.0 cycles per second)⁵⁰. Voltage-dependent membrane properties such as those in STG neurons are common features of all CPG neurons, although in many cases their specific roles within these circuits are yet to be elucidated^{2,3}.

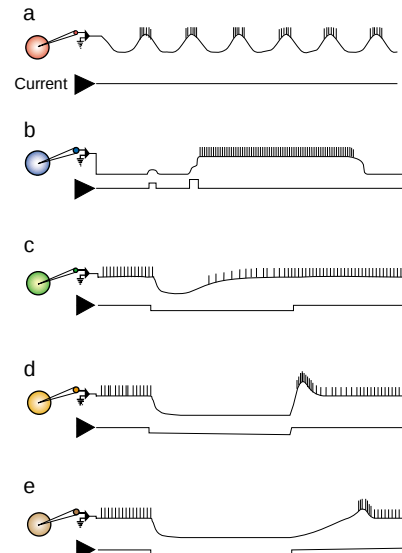


Figure 2 Illustrations of some of the membrane properties occurring in intracellularly recorded neurons (coloured circles) of central pattern generating circuits, including those in the STG. Each current trace is a monitor of depolarizing (up) and/or hyperpolarizing (down) current injections into the recorded neuron. Comparable events can be elicited by excitatory and inhibitory synaptic input. **a**, An endogenous oscillator undergoes rhythmic membrane potential oscillations without requiring any rhythmic input. **b**, The plateau potential is a persistent depolarizing response that generates persistent action potential activity and outlasts the triggering stimulus. It has a membrane potential threshold for its activation and it eventually self-terminates. **c**, Escape from inhibition is an excitation that is triggered by a sufficient amplitude and duration of hyperpolarization. It enables the neuron to depolarize, often to the point of firing action potentials, despite the continued presence of a hyperpolarizing input. **d**, Post-inhibitory rebound (PIR) is also triggered by hyperpolarization, but it represents an overshooting of the resting potential with an associated burst of action potentials after termination of the hyperpolarization. **e**, PIR delay shows a slower rebound depolarization, compared with PIR, after a period of hyperpolarization.

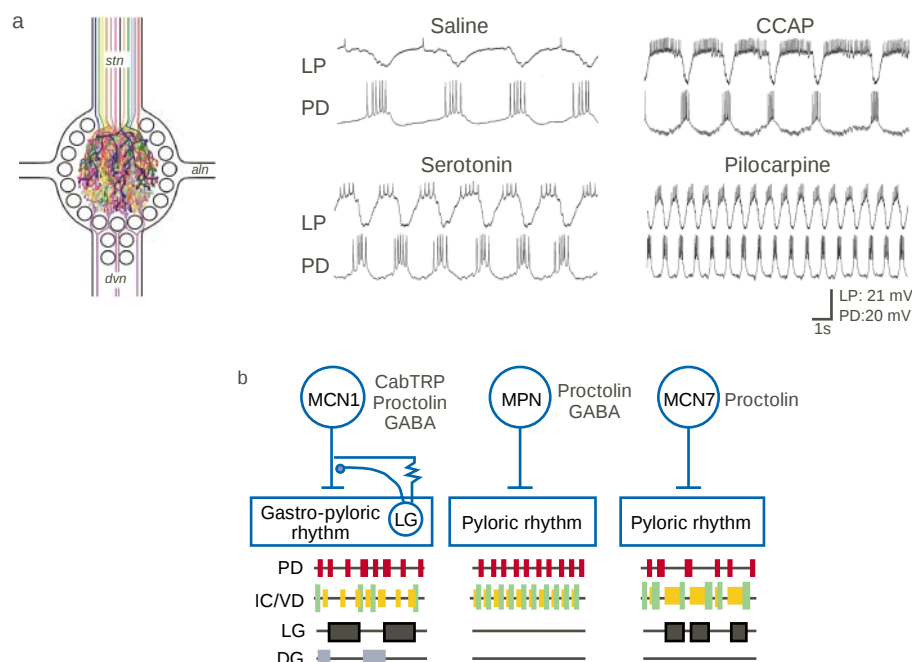
Some non-CPG circuits also combine synaptic inhibition and membrane properties to generate patterned neuronal output. One well-characterized example is the circuit underlying sleep-related rhythms in the thalamus⁶. In this system, the basic circuit includes a spatially iterated set of reciprocally connected inhibitory (perigeniculate/reticular) neurons and excitatory (thalamocortical) neurons. The activity pattern generated by these neurons relies on the effective activation of such membrane properties as post-inhibitory rebound, escape from inhibition and plateau-like potentials. The switch from the sleep to the waking state in this system seems to result largely from modulatory inputs that depolarize the circuit neurons sufficiently to prevent the same inhibitory synapses within the circuit from activating these membrane properties⁶.

Each of the membrane properties in STG neurons is generally associated with a particular set of ionic currents, although in many cases more than one set of currents can mediate the same event⁵¹. Molecular approaches have provided a more detailed description of some of the channels and currents underlying these properties, including the genes that encode them and their specific location(s) within individual neurons^{52,53}. This work has also been extended to the functional level using multiphoton imaging and voltage-clamp analysis of Ca^{2+} currents in an identified STG neuron⁵⁴.

These studies provide the basis for a more accurate understanding of the physiological properties of individual circuit neurons, as well as their response to synaptic and modulatory inputs. It has become evident that mammalian dendrites also exhibit regions with distinct

Figure 3 Modulation of the pyloric and gastric mill rhythms in the STG.

a. Schematic of the modulatory innervation of the STG (left panel; adapted from ref. 4). Each colour represents a distinct neuromodulatory transmitter or complement of transmitters localized to the axon and STG terminals of projection neurons that innervate the STG via the stomatogastric nerve (*stn*) and of sensory neurons that innervate the STG via the dorsal ventricular nerve (*dvn*). Individual bath application of different modulators to the isolated STG in the crab *Cancer borealis* elicits different versions of the pyloric rhythm (right panel; S. R. Hertzberg, M.P.B. & M.P.N., unpublished data). The pyloric rhythm is monitored by intracellular recordings of two pyloric circuit neurons (PD and LP). Modulator concentrations: serotonin (10^{-5} M), crustacean cardioactive peptide (CCAP; 10^{-6} M), and the muscarinic agonist pilocarpine (10^{-5} M). Each application



was followed by a saline wash, during which the rhythm returned to control conditions (saline). All recordings are from the same preparation. **b.** Selective activation of distinct modulatory projection neurons which have a peptide co-transmitter in common (proctolin) have different actions on the pyloric and gastric mill circuits in the STG. Note that the inhibitory and electrical synapses from the LG neuron to MCN1 occur within the STG neuropil^{29,71}. Abbreviations: CabTRP, *Cancer borealis* tachykinin-related peptide; MCN1/7, modulatory commissural neuron 1/7; MPN, modulatory proctolin neuron. Modified from ref. 11.

physiological properties, as a result of clustering of distinct channels⁵⁵. Additionally, the molecular characterization of the ion channels and currents present in STG neurons is facilitating a more accurate determination of the identity or distinctiveness of similar neurons. The use of molecular identity should be particularly valuable in more complex systems where there exist populations of apparently equivalent neurons.

At most of the STG circuit synapses, graded transmitter release is predominant¹⁶. This occurs independently of action potentials and, once past a threshold level, increases in a graded fashion with increasing depolarization. The mechanisms underlying this release, as well as its modulation, have been studied extensively in the STG^{16,56}, where important principles for comparable signalling events in other systems have been generated. Although graded release is pivotal within the STG circuit, these same circuit neurons use action potentials for exporting the STG circuit output to the muscles and to more central ganglia.

Modelling of circuit activity has become a valuable tool for elucidating neural circuit operation, even in small systems such as the STG, because of the complexity of neuronal circuits relative to available experimental techniques. This approach, combined with experimental tests of the predictions from each model, has been instrumental in determining the function of a number of specific cellular and synaptic properties of STG circuit neurons^{57–60}. For example, efforts have led to several non-intuitive insights, including the presence of distinct types of endogenous oscillatory properties⁶¹, the use of membrane properties to transduce temporal patterns into a neural code⁵⁸, and the presence of intercircuit coordination^{20,62}.

Electrical coupling

Gap junction-mediated electrical coupling is prevalent throughout both vertebrate and invertebrate nervous systems. It links functionally equivalent neurons into co-active groups, provides fast and reliable synaptic communication and can enable coincidence detection^{63–66}. Electrical coupling also provides additional degrees of freedom

within neuronal circuits^{67,68}. For example, within the pyloric circuit there are several neuronal pairs that are connected synaptically via both chemical and electrical synapses (Fig. 1b). Under steady-state conditions, inhibition is the dominant action at these dual synapses and the net result is that the coupled neurons fire in alternation instead of synchronously^{69,70}. However, in certain modulatory environments the relative strength of these two types of synapses changes, causing a switch in the net response from inhibition to excitation⁷⁰.

Another unusual use of electrical synapses occurs in the gastric mill circuit, where there is a voltage-dependent electrical synapse between a gastric mill neuron (the lateral gastric neuron) and the STG terminals of a modulatory projection neuron called MCN1 (ref. 71; and Fig. 3b). Because of the voltage-dependence of the electrical synapse and the particular circuit configuration, the projection neuron excites the gastric mill circuit via its released co-transmitters during one half of each cycle and via electrical coupling during the other half of each cycle⁷¹.

Even when neuronal ensembles are tightly electrically coupled, as occurs among the pyloric pacemaker group, the different components can have different functional roles^{17,18}. In the spiny lobster, the pyloric pacemaker group includes a single oscillator neuron (called the anterior burster neuron), plus paired pyloric dilator neurons. Their electrical coupling enables the slow oscillations of their membrane potentials to occur synchronously, and both neuron types make inhibitory synapses on nearly all other pyloric neurons. However, they do so using different neurotransmitters⁶⁷ (Fig. 1b). The glutamatergic inhibition from the anterior burster neuron has a relatively fast onset and offset, whereas the cholinergic pyloric dilator neuron evokes a slower onset and longer lasting inhibition in the same target neurons. One way by which modulatory inputs elicit different pyloric rhythms is by having differential actions on these two pacemaker neurons^{17,18,56,67}. For example, dopamine application enhances the activity of the anterior burster neuron while inhibiting the pyloric dilator neuron^{18,67}. In this case, the faster synaptic actions of the anterior burster neuron are dominant, contributing to a

change in the timing of activity in the other pyloric neurons. These actions of dopamine work in concert with dopamine modulation of a specific potassium current (I_A) in the other pyloric neurons to bring about the altered timing of their activity⁵⁶. The extent to which different subsets of a population of electrically coupled neurons have different roles in circuit activity remains to be determined in most systems.

Another functional distinction between the electrically coupled anterior burster and pyloric dilator neurons was discovered from determining how their individual responses to neuromodulators influences the pyloric cycle frequency^{17,18}. For example, the neuropeptide proctolin increases the speed of weakly cycling pyloric rhythms, but it evokes a maximal pyloric cycle frequency of 1 cycle s^{-1} despite the ability of this rhythm to cycle faster under other conditions. By studying the actions of this peptide separately on each pyloric circuit neuron, Hooper and Marder¹⁷ found that proctolin increased the rhythmic cycling of the isolated anterior burster neuron to frequencies of 2–3 cycles s^{-1} , but it had no effect on the isolated pyloric dilator neurons. By sequentially photoinactivating each pyloric dilator neuron in the otherwise intact circuit, they showed that the lack of responsiveness of these neurons to proctolin slows the cycling speed of the anterior burster neuron, via their electrical coupling. This was a clear indication that the non-responsiveness of a neuron to a neuromodulator can have an impact on the circuit response to that modulator, and highlights the importance of knowing the function(s) of all circuit components for fully understanding circuit dynamics.

Short-term synaptic dynamics

Short-term synaptic dynamics include processes, such as short-term depression and facilitation, that have been studied extensively at the level of individual synapses in many systems⁷². The contribution of this form of synaptic plasticity to circuit dynamics has been intimated from studies of cortical neurons⁷², but a direct role for these processes within an operating circuit has only recently been documented. For example, modulation of the spinal locomotor CPG in lamprey alters the impact of synaptic depression and facilitation on the CPG output⁷³. In the stomatogastric system, regulation of synaptic depression seems to be a determining factor for whether the pyloric cycle frequency is controlled by the endogenous oscillatory property of the pyloric pacemaker neuron or the inhibitory synaptic input that this neuron receives from the lateral pyloric neuron^{72,74} (Fig. 1b). Under steady-state conditions, the latter synapse is effectively depressed and the intrinsic properties of the pacemaker neuron control pyloric cycle speed. However, it seems likely that under some modulatory conditions the lateral pyloric synapse onto the pacemakers is strengthened such that it effectively regulates the speed of this rhythm. In another study, Combes *et al.*³⁰ showed that short-term facilitation is used to switch the version of the gastric mill rhythm that is elicited by activation of an identified sensory pathway. When the sensory neuron fires at a moderate level, it preferentially activates one particular projection neuron that drives the gastric mill rhythm. However, higher-frequency firing of the sensory neuron produces facilitating excitatory postsynaptic potentials in another projection neuron whose activation changes the gastric mill rhythm.

Activity-dependent plasticity

Considerable attention has focused on homeostatic regulation of neural network activity^{75,76}. Some of the pioneering studies of this phenomenon, performed with lobster STG neurons⁷⁷, documented the ability of neurons to compensate for long-lasting changes in their activity by altering their cellular properties, with the consequence that their activity level ultimately returned to its previous steady state. These compensatory changes in the activity of pyloric neurons are mediated by long-term changes in the balance of active ionic currents^{57,78}. One remarkable conclusion from this work has been that the same neuronal activity pattern can be achieved in the same

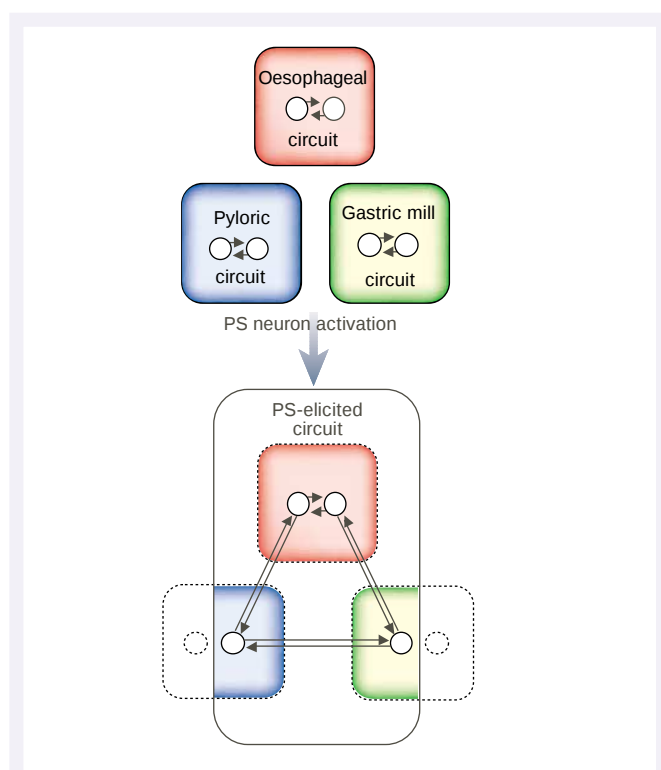


Figure 4 Modulatory input can dismantle and reconfigure the STG circuits. Generally, the gastric mill, pyloric and oesophageal circuits generate distinct motor rhythms. However, activation of some modulatory inputs, such as the pyloric suppressor (PS) neuron, can eliminate these rhythms and replace them with a single, conjoint motor pattern. The PS-elicited motor circuit includes a subset of the neurons (circles within coloured regions) that comprise the three circuits. The PS neuron also inhibits the activity of the gastric mill and pyloric circuit neurons (circles with dashed outlines located outside coloured regions) that do not participate in this conjoint motor pattern. Adapted from ref. 83; see also ref. 46.

neuron by different balances of ionic currents. Previous studies involving modulation of individual ion currents focused on the ability of this modulation to evoke short-term changes in the activity pattern of the neuron. Clearly, this same event can preserve the original activity pattern of the neuron over the long term. Additionally, because the STG work was performed in a network context, we learned that not only is the individual neuronal activity conserved, but also the circuit output.

Intercircuit coordination

Many behaviours involve the coordination of different sets of movements, such as the coordination of breathing with strenuous activities like running or swimming. There are few existing cellular-level biological models for how the distinct neuronal circuits underlying these movements coordinate their activity patterns. This type of coordination is being elucidated using the gastric mill and pyloric circuits. These two circuits operate in distinct time domains, with the pyloric rhythm ranging from ~0.5–2.0 cycles s^{-1} and the gastric mill rhythm ranging from ~0.05–0.2 cycles s^{-1} (Fig. 1a), and they can operate independently^{20,79}. There are, however, identified synapses that enable each circuit to regulate the output of the other^{20,79}. The gastric mill circuit regulates the pyloric rhythm through its ability to synaptically inhibit the STG terminals of the modulatory projection neuron MCN1, whose activity drives the gastric mill rhythm and enhances the pyloric rhythm^{29,71,79} (Fig. 3b). Thus, during one half of each gastric mill cycle, the lateral gastric neuron inhibits MCN1 and thereby eliminates the MCN1 excitation of the pyloric circuit. During the other half of each cycle the lateral

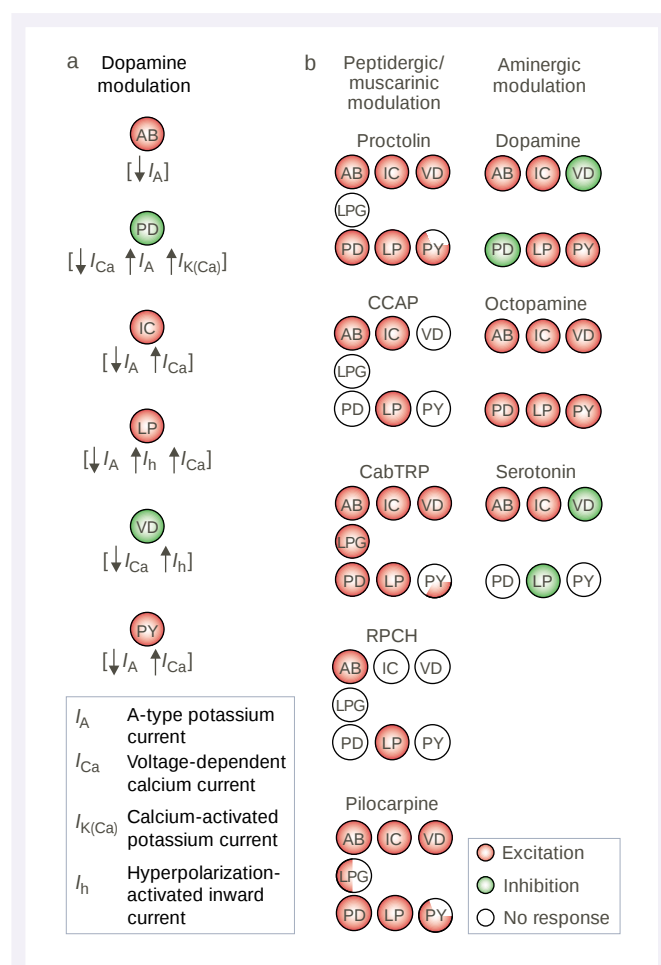


Figure 5 Convergence and divergence of transmitter actions. **a**, Dopamine influences a distinct but overlapping set of ionic currents in each pyloric circuit neuron. Moreover, it can have opposite effects on the same current in different neurons. For example, dopamine has opposite effects on I_h in the electrically coupled AB and PD neurons. Adapted from ref. 56. **b**, Different neuromodulators influence overlapping but distinct subsets of pyloric circuit neurons. All of the peptidergic and muscarinic actions are excitatory, as are the actions of octopamine. Dopamine and serotonin excite some pyloric neurons and inhibit others. RPCH, red pigment concentrating hormone; LPG, lateral posterior gastric neuron. Peptides and muscarinic actions adapted from ref. 91; amines adapted from ref. 15.

gastric neuron is silent, enabling MCN1 to release its co-transmitters and excite the pyloric circuit.

Insight into the pyloric circuit regulation of the gastric mill rhythm developed from a computational model of the gastric mill rhythm elicited by MCN1 stimulation⁶². By manipulating the biological circuits, Bartos *et al.*²⁰ verified the predictions of the model that the pyloric circuit regulates the speed of the gastric mill rhythm via an identified synapse, and that this same synapse links the start of each gastric mill cycle to the start of a pyloric cycle, thereby coordinating their activity patterns. There are comparable examples of coordination between distinct motor patterns in numerous systems, but the cellular-level mechanisms underlying this coordination are yet to be determined^{80,81}.

Neuromodulation in the stomatogastric system

Early work on neural circuit modulation in the isolated STG showed that bath-applied neuromodulators reproducibly elicit stable, distinct pyloric and gastric mill rhythms, enabling a detailed analysis of these events⁴⁷. The discovery that CPGs are functionally flexible marked a paradigm shift that was subsequently extended to all types

of neuronal circuits. This ability to modify circuit output is largely a consequence of the relatively slow time course of modulatory actions and the fact that their primary targets are voltage- and time-dependent currents⁸². These currents, by definition, are each active in only a limited range of membrane potentials outside of which they have no influence. Comparable application of transmitters with ionotropic actions instead drives all target neurons towards the reversal potential of the current(s) opened by the transmitter and thereby disrupts any ongoing rhythmic activity⁸².

The degree of flexibility afforded to neuronal networks by modulatory inputs is extensive. For example, the same STG circuit neuron can be active with the pyloric rhythm, gastric mill rhythm or both rhythms simultaneously¹³. The same circuit neuron can also be switched to participate in different rhythms^{46,83,84}. The gastric mill and pyloric circuits can even be dismantled, with some components participating in a different motor pattern and other components remaining silent^{46,83} (Fig. 4). It has not yet been possible to prove that comparable events occur in the vertebrate CNS, but the available evidence is strongly suggestive^{85–87}.

One realization that developed from these studies of neuromodulation is that the activity phenotype, membrane properties and synaptic actions of individual neurons are state dependent. For example, the same neuron under different conditions can be an endogenous oscillator, generate plateau potentials and/or post-inhibitory rebound, or express none of these properties². Similarly, the strength, sign or even presence of synaptic actions can come and go. Changes such as these have significant consequences for circuit output^{2,56,88}.

Modulation of CPG activity in the STG involves many different small-molecule, peptide and gaseous neurotransmitters^{8,35,89}. The second-messenger pathways involved in these modulatory actions are not well described, but it is clear from imaging studies in physiological preparations that cyclic AMP levels within the STG are upregulated in unique ways by different neuromodulators⁹⁰.

The presence of so many different neuromodulators in this system has provided an opportunity to determine the extent to which there is convergence and divergence of action. The actions produced by muscarinic acetylcholine, biogenic amines and neuropeptides in the STG provide examples of both^{21,56,91}. For example, dopamine has a different effect on each pyloric neuron owing to its differential actions on multiple ion currents⁵⁶ (Fig. 5a). In contrast, at least four different neuropeptides plus muscarinic agonists have convergent actions on a single, voltage-dependent ion current in the pyloric neurons^{56,91} (Fig. 5b). Despite this convergence at the current level, each of the modulators belonging to this latter group elicits different pyloric rhythms^{47,92} (Fig. 3a). At least part of this apparent discrepancy is because each modulator has direct actions on a distinct but overlapping subset of pyloric neurons⁹¹ (Fig. 5b). The same is true for the biogenic amines dopamine, serotonin and octopamine¹⁵ (Fig. 5b). In fact, the pyloric rhythm elicited by crustacean cardioactive peptide can be transformed into the pyloric rhythm elicited by the peptide proctolin by injecting the dynamic clamp version of the peptide-activated current into two pyloric neurons that respond only to proctolin²¹. Clearly, there is not a single underlying principle to describe the means by which different modulators elicit distinct outputs from the same network.

Co-transmission

In most systems, modulation of neural circuit activity is studied by direct application of a neuromodulator, usually because it is difficult to identify and selectively activate the modulatory neurons that release these substances. However, CPGs can generate distinct motor patterns not only when different modulators are applied directly to the isolated nervous system (Fig. 3a), but also when different modulatory neurons are activated^{11,15,36,93} (Fig. 3b).

Some early reports from work in the stomatogastric system supported the hypothesis that bath application of neuromodulators is equivalent to activating the relevant modulatory neuron^{11,15,34}.

However, modulatory neurons in this and other systems also utilize mechanisms that cannot be replicated using bath-application techniques^{11,46,83,94}. For example, neurons commonly contain more than one neurotransmitter: often at least one neuropeptide and a small-molecule transmitter^{11,35}. In addition to modulatory actions, the small-molecule transmitter usually has ionotropic actions, and ionotropic actions on rhythmically active circuits are not effectively mimicked by bath application.

Several modulatory neurons with identified co-transmitters occur in the stomatogastric system, including projection neurons¹¹ (Fig. 3b) and sensory neurons^{8,34}. Work with these neurons has illustrated that their co-transmitters exhibit both convergence and divergence of action. In some cases, these projection neurons affect subsets of their targets via only some of their co-transmitters^{11,95}. Additionally, different projection neurons can elicit distinct STG rhythms despite having a neuropeptide (proctolin) transmitter in common¹¹ (Fig. 3b). This results in part from the presence of distinct co-transmitters in these neurons, but also from a difference in the strength of the proctolin action onto the same pyloric neurons when this peptide is released from different projection neurons^{11,96}. This latter result is at least partly a consequence of a differential regulation of proctolin, released from each neuron, by extracellular peptidase activity⁹⁶.

Development of circuit modulation

Neural circuit development has been studied extensively in the vertebrate CNS, although little is known regarding the physiological development of neural circuits at the cellular level. However, recent developmental studies in the stomatogastric system are providing information that resonates with and extends those findings in the vertebrate systems^{97,98}. For example, prior to the onset of feeding behaviour in the embryonic lobster, the STG rhythms are slower and less regular than in the adult, a situation paralleled in other developing networks^{99,100}. Furthermore, the STG circuits are rhythmically active before the complete maturation of their modulatory inputs. This results in distinct activity patterns at different developmental times. Gradual maturation of the modulatory projection neurons to the STG through development seems to result primarily from the distinct developmental time at which different neuromodulators, including co-localized ones, first become expressed.

Closing the gaps

So far, most attention in the STG and comparable systems has focused on relating cellular and synaptic properties to circuit dynamics under steady-state and modulatory conditions. The stage is now set for expanding the scope of this endeavour to focus more intensively at how events at both the molecular and systems levels contribute to the control of neuronal circuit output. Additionally, to better bridge the gap between *in vitro* studies and normal activity in the intact animal, the next step is likely to involve working with the stomatogastric system while it remains connected with the rest of the CNS, isolated from the rest of the animal.

Previously, the complexity within invertebrate neural systems was thought to be a specialized adaptation of these animals that enabled them to use a relatively small number of neurons to perform a rich repertoire of behaviours. Because of their significantly larger numbers, neurons in the vertebrate nervous system were not expected to have the same range of computational complexity. It was therefore not certain whether the same basic principles guided neural circuit operation in these two groups of animals. It is now evident that vertebrate neurons and neural circuits are as multifunctional as their invertebrate counterparts^{2,3,6,87}. Because of the similarities underlying neural circuit activity in both groups of animals, invertebrate models such as the crustacean stomatogastric nervous system continue to provide considerable insight into how the comparable neural circuits operate in the numerically larger and less accessible vertebrate CNS. □

function in the nervous system. *J. Physiol.* **48**, 18–46 (1914).

2. Marder, E. & Calabrese, R. L. Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
3. Stein, P. S. G., Grillner, S., Selverston, A. I. & Stuart, D. G. (eds) *Neurons, Networks, and Motor Behavior* (The MIT Press, Cambridge, MA, 1997).
4. Marder, E. & Bucher, D. Central pattern generators and the control of rhythmic movements. *Curr. Biol.* **11**, R986–R996 (2001).
5. Pearson, K. G. Neural adaptation in the generation of rhythmic behavior. *Annu. Rev. Physiol.* **62**, 723–753 (2000).
6. McCormick, D. A. & Bal, T. Sleep and arousal: thalamocortical mechanisms. *Annu. Rev. Neurosci.* **20**, 185–215 (1997).
7. Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M. (eds) *Dynamic Biological Networks: The Stomatogastric Nervous System* (MIT Press, Cambridge, MA, 1992).
8. Skieba, P. Neuropeptides are ubiquitous chemical mediators: using the stomatogastric nervous system as a model system. *J. Exp. Biol.* **204**, 2035–2048 (2001).
9. Turrigiano, G. G. & Heinzel, H.-G. in *Dynamic Biological Networks: The Stomatogastric Nervous System* (eds Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M.) 197–220 (MIT Press, Cambridge, MA, 1992).
10. Clemens, S., Combes, D., Meyrand, P. & Simmers, J. Long-term expression of two interacting motor pattern-generating networks in the stomatogastric system of freely behaving lobster. *J. Neurophysiol.* **79**, 1396–1408 (1998).
11. Nusbaum, M. P., Blitz, D. M., Swensen, A. M., Wood, D. & Marder, E. The roles of co-transmission in neural network modulation. *Trends Neurosci.* **24**, 146–154 (2001).
12. Heinzel, H. G., Weimann, J. M. & Marder, E. The behavioral repertoire of the gastric mill in the crab, *Cancer pagurus*: an in situ endoscopic and electrophysiological examination. *J. Neurosci.* **13**, 1793–1803 (1993).
13. Weimann, J. M. & Marder, E. Switching neurons are integral members of multiple oscillatory networks. *Curr. Biol.* **4**, 896–902 (1994).
14. Miller, J. P. in *The Crustacean Stomatogastric System* (eds Selverston, A. I. & Moulins, M.) 109–136 (Springer, Berlin, 1987).
15. Harris-Warrick, R. M., Nagy, F. & Nusbaum, M. P. in *Dynamic Biological Networks: The Stomatogastric Nervous System* (eds Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M.) 87–138 (MIT Press, Cambridge, MA, 1992).
16. Hartline, D. K. & Graubard, K. in *Dynamic Biological Networks: The Stomatogastric Nervous System* (eds Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M.) 31–86 (MIT Press, Cambridge, MA, 1992).
17. Hooper, S. L. & Marder, E. Modulation of the lobster pyloric rhythm by the peptide proctolin. *J. Neurosci.* **7**, 2097–2112 (1987).
18. Ayali, A. & Harris-Warrick, R. M. Monoamine control of the pacemaker kernel and cycle frequency in the lobster pyloric network. *J. Neurosci.* **19**, 6712–6722 (1999).
19. Sharp, A. A., O'Neil, M. B., Abbott, L. F. & Marder, E. The dynamic clamp: artificial conductances in biological neurons. *Trends Neurosci.* **16**, 389–394 (1993).
20. Bartos, M., Manor, Y., Nadim, F., Marder, E. & Nusbaum, M. P. Coordination of fast and slow rhythmic neuronal circuits. *J. Neurosci.* **19**, 6650–6660 (1999).
21. Swensen, A. M. & Marder, E. Modulators with convergent cellular actions elicit distinct circuit outputs. *J. Neurosci.* **21**, 4050–4058 (2001).
22. Selverston, A. I., Kleindienst, H. U. & Huber, F. Synaptic connectivity between cricket auditory interneurons as studied by selective photoinactivation. *J. Neurosci.* **5**, 1283–1292 (1985).
23. Warzecha, A. K., Egelhaaf, M. & Borst, A. Neural circuit tuning fly visual interneurons to motion of small objects. I. Dissection of the circuit by pharmacological and photoinactivation techniques. *J. Neurophysiol.* **69**, 329–339 (1993).
24. Kemenes, G. & Elliott, C. J. Analysis of the feeding motor pattern in the pond snail, *Lymnaea stagnalis*: photoinactivation of axonally stained pattern-generating interneurons. *J. Neurosci.* **14**, 153–166 (1994).
25. Ulrich, D. & Huguenard, J. R. GABA_A-receptor-mediated rebound burst firing and burst shunting in thalamus. *J. Neurophysiol.* **78**, 1748–1751 (1997).
26. Jaeger, D. & Bower, J. Synaptic control of spiking in cerebellar Purkinje cells: dynamic current clamp based on model conductances. *J. Neurosci.* **19**, 6090–6101 (1999).
27. Kiehn, O., Kjaerulff, O., Tresch, M. C. & Harris-Warrick, R. Contributions of intrinsic motor neuron properties to the production of rhythmic motor output in the mammalian spinal cord. *Brain Res.* **53**, 649–659 (2000).
28. Elson, R. C. & Selverston, A. I. Mechanisms of gastric rhythm generation in the isolated stomatogastric ganglion of spiny lobsters: bursting pacemaker potentials, synaptic interactions, and muscarinic modulation. *J. Neurophysiol.* **68**, 890–907 (1992).
29. Coleman, M. J. & Nusbaum, M. P. Functional consequences of compartmentalization of synaptic input. *J. Neurosci.* **14**, 6544–6552 (1994).
30. Combes, D., Meyrand, P. & Simmers, J. Dynamic restructuring of a rhythmic motor program by a single mechanoreceptor neuron in lobster. *J. Neurosci.* **19**, 3620–3628 (1999).
31. Koshiya, N. & Smith, J. C. Neuronal pacemaker for breathing visualized *in vitro*. *Nature* **400**, 360–363 (1999).
32. Parker, D. & Grillner, S. Neuronal mechanisms of synaptic and network plasticity in the lamprey spinal cord. *Prog. Brain Res.* **125**, 381–398 (2000).
33. Thoby-Brisson, M. & Ramirez, J. M. Role of inspiratory pacemaker neurons in mediating the hypoxic response of the respiratory network *in vitro*. *J. Neurosci.* **20**, 5858–5866 (2000).
34. Katz, P. S. & Harris-Warrick, R. M. Neuromodulation of the crab pyloric central pattern generator by serotonergic/cholinergic proprioceptive afferents. *J. Neurosci.* **10**, 1495–1512 (1990).
35. Marder, E., Christie, A. E. & Kilman, V. L. Functional organization of cotransmission systems: lessons from small nervous systems. *Invert. Neurosci.* **1**, 105–112 (1995).
36. Norris, B. J., Coleman, M. J. & Nusbaum, M. P. Pyloric motor pattern modification by a newly identified projection neuron in the crab stomatogastric nervous system. *J. Neurophysiol.* **75**, 97–108 (1996).
37. Combes, D., Meyrand, P. & Simmers, J. Motor pattern specification by dual descending pathways to a lobster rhythm-generating network. *J. Neurosci.* **19**, 3610–3619 (1999).
38. Morris, L. G., Thuma, J. B. & Hooper, S. L. Muscles express motor patterns of non-innervating neural networks by filtering broad-band input. *Nature Neurosci.* **3**, 245–250 (2000).
39. Hooper, S. L. & Weaver, A. L. Motor neuron activity is often insufficient to predict motor response.

1. Brown, T. G. On the nature of the fundamental activity of the nervous centres; together with an analysis of the conditioning of rhythmic activity in progression, and a theory of the evolution of

- Curr. Opin. Neurobiol.* **10**, 676–682 (2000).
40. Katz, P. S., Kirk, M. D. & Govind, C. K. Facilitation and depression at different branches of the same motor axon: evidence for presynaptic differences in release. *J. Neurosci.* **13**, 3075–3089 (1993).
 41. Jorge-Rivera, J. C., Sen, K., Birmingham, J. T., Abbott, L. F. & Marder, E. Temporal dynamics of convergent modulation at a crustacean neuromuscular junction. *J. Neurophysiol.* **80**, 2559–2570 (1998).
 42. Gutovitz, S., Birmingham, J. T., Luther, J. A., Simon, D. J. & Marder, E. GABA enhances transmission at an excitatory glutamatergic synapse. *J. Neurosci.* **21**, 5935–5943 (2001).
 43. Katz, P. S. & Tazaki, K. in *Dynamic Biological Networks: The Stomatogastric Nervous System* (eds Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M.) 221–262 (MIT Press, Cambridge, MA, 1992).
 44. Katz, P. S. & Harris-Warrick, R. M. The evolution of neuronal circuits underlying species-specific behavior. *Curr. Opin. Neurobiol.* **9**, 628–633 (1999).
 45. Meyrand, P., Faumont, S., Simmers, J., Christie, A. E. & Nusbaum, M. P. Species-specific modulation of pattern-generating circuits. *Eur. J. Neurosci.* **12**, 2585–2596 (2000).
 46. Meyrand, P., Simmers, J. & Moulins, M. Dynamic construction of a neural network from multiple pattern generators in the lobster stomatogastric nervous system. *J. Neurosci.* **14**, 630–644 (1994).
 47. Marder, E. & Hooper, S. L. in *Model Neural Networks and Behavior* (ed. Selverston, A. I.) 319–337 (Plenum, New York, 1985).
 48. Tierney, A. J. & Harris-Warrick, R. M. Physiological role of the transient potassium current in the pyloric circuit of the lobster stomatogastric ganglion. *J. Neurophysiol.* **67**, 599–609 (1992).
 49. Elson, R. C., Huerta, R., Abarbanel, H. D., Rabinovich, M. I. & Selverston, A. I. Dynamic control of irregular bursting in an identified neuron of an oscillatory circuit. *J. Neurophysiol.* **82**, 115–122 (1999).
 50. Hooper, S. L. Phase maintenance in the pyloric pattern of the lobster (*Panulirus interruptus*) stomatogastric ganglion. *J. Comput. Neurosci.* **4**, 191–205 (1997).
 51. Harris-Warrick, R. M. & Flamm, R. E. Multiple mechanisms of bursting in a conditional bursting neuron. *J. Neurosci.* **7**, 2113–2128 (1987).
 52. Baro, D. J. & Harris-Warrick, R. M. Differential expression and targeting of K⁺ channel genes in the lobster pyloric central pattern generator. *Ann. NY Acad. Sci.* **860**, 281–295 (1998).
 53. Baro, D. J. et al. Molecular underpinnings of motor pattern generation: differential targeting of shal and shaker in the pyloric motor system. *J. Neurosci.* **20**, 6619–6630 (2000).
 54. Kloppenburg, P., Zipfel, W. R., Webb, W. W. & Harris-Warrick, R. M. Highly localized Ca²⁺ accumulation revealed by multiphoton microscopy in an identified motoneuron and its modulation by dopamine. *J. Neurosci.* **20**, 2523–2533 (2000).
 55. Reyes, A. Influence of dendritic conductances on the input-output properties of neurons. *Annu. Rev. Neurosci.* **24**, 653–675 (2001).
 56. Harris-Warrick, R. M. et al. Distributed effects of dopamine modulation in the crustacean pyloric network. *Ann. NY Acad. Sci.* **860**, 155–167 (1998).
 57. Goldman, M. S., Golowasch, J., Marder, E. & Abbott, L. F. Global structure, robustness, and modulation of neuronal models. *J. Neurosci.* **21**, 5229–5238 (2001).
 58. Hooper, S. L. Transduction of temporal patterns by single neurons. *Nature Neurosci.* **1**, 720–726 (1998).
 59. Soto-Treviño, C., Thoroughman, K. A., Marder, E. & Abbott, L. F. Activity-dependent modification of inhibitory synapses in models of rhythmic neural networks. *Nature Neurosci.* **4**, 297–303 (2000).
 60. Selverston, A. I. et al. Reliable circuits from irregular neurons: a dynamical approach to understanding central pattern generators. *J. Physiol. Paris* **94**, 357–374 (2000).
 61. Marder, E. et al. Physiological insights from cellular and network models of the stomatogastric nervous systems of lobsters and crabs. *Am. Zool.* **33**, 29–39 (1993).
 62. Nadim, F., Manor, Y., Nusbaum, M. P. & Marder, E. Frequency regulation of a slow rhythm by a fast periodic input. *J. Neurosci.* **18**, 5053–5067 (1998).
 63. Edwards, D. H., Yeh, S. R. & Krasne, F. B. Neuronal coincidence detection by voltage-sensitive electrical synapses. *Proc. Natl Acad. Sci. USA* **95**, 7145–7150 (1998).
 64. Perez-Velazquez, J. L. & Carlen, P. L. Gap junctions, synchrony and seizures. *Trends Neurosci.* **23**, 68–74 (2000).
 65. Galarreta, M. & Hestrin, S. Electrical synapses between GABA-releasing interneurons. *Nature Rev. Neurosci.* **2**, 425–433 (2001).
 66. Schmitz, D. et al. Axo-axonal coupling: a novel mechanism for ultrafast neuronal communication. *Neuron* **31**, 831–840 (2001).
 67. Marder, E. Roles for electrical coupling in neural circuits as revealed by selective neuronal deletions. *J. Exp. Biol.* **112**, 147–167 (1984).
 68. Kiehn, O. & Tresch, M. C. Gap junctions and motor behavior. *Trends Neurosci.* **25**, 108–115 (2002).
 69. Graubard, K. & Hartline, D. K. Full-wave rectification from a mixed electrical-chemical synapse. *Science* **237**, 535–537 (1987).
 70. Johnson, B. R., Peck, J. H. & Harris-Warrick, R. M. Differential modulation of chemical and electrical components of mixed synapses in the lobster stomatogastric ganglion. *J. Comp. Physiol. A* **175**, 233–249 (1994).
 71. Coleman, M. J., Meyrand, P. & Nusbaum, M. P. A switch between two modes of synaptic transmission mediated by presynaptic inhibition. *Nature* **378**, 502–505 (1995).
 72. Nadim, F. & Manor, Y. The role of short-term synaptic dynamics in motor control. *Curr. Opin. Neurobiol.* **10**, 683–690 (2000).
 73. Parker, D. & Grillner, S. Activity-dependent metaplasticity of inhibitory and excitatory synaptic transmission in the lamprey spinal cord locomotor network. *J. Neurosci.* **19**, 1647–1656 (1999).
 74. Manor, Y. & Nadim, F. Synaptic depression mediates bistability in neuronal networks with recurrent inhibitory connectivity. *J. Neurosci.* **21**, 9460–9470 (2001).
 75. Turrigiano, G. Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* **22**, 221–217 (1999).
 76. Turrigiano, G. G. & Nelson, S. B. Hebb and homeostasis in neuronal plasticity. *Curr. Opin. Neurobiol.* **10**, 358–364 (2000).
 77. Turrigiano, G. G., LeMasson, G. & Marder, E. Selective regulation of current densities underlies spontaneous changes in the activity of cultured neurons. *J. Neurosci.* **15**, 3640–3652 (1995).
 78. Mizrahi, A. et al. Long-term maintenance of channel distribution in a central pattern generator neuron by neuromodulatory inputs revealed by decentralization in organ culture. *J. Neurosci.* **21**, 7331–7339 (2001).
 79. Bartos, M. & Nusbaum, M. P. Intercircuit control of motor pattern modulation by presynaptic inhibition. *J. Neurosci.* **17**, 2247–2256 (1997).
 80. Kawahara, K., Kumagai, S., Nakazono, Y. & Myamoto, Y. Coupling between respiratory and stepping rhythms during locomotion in decerebrate cats. *J. Appl. Physiol.* **67**, 110–115 (1989).
 81. Larson, C. R., Yajima, Y. & Ko, P. Modification of activity of medullary respiratory-related neurons for vocalizing and swallowing. *J. Neurophysiol.* **71**, 2294–2304 (1994).
 82. Marder, E. & Meyrand, M. in *Neuronal and Cellular Oscillators* (ed. Jacklet, J.) 317–338 (Dekker, New York, 1989).
 83. Meyrand, P., Simmers, J. & Moulins, M. Construction of a pattern-generating circuit with neurons of different networks. *Nature* **351**, 60–63 (1991).
 84. Hooper, S. L. & Moulins, M. Cellular and synaptic mechanisms responsible for a long-lasting restructuring of the lobster pyloric network. *J. Neurophysiol.* **64**, 1574–1589 (1990).
 85. Soffe, S. R. Two distinct rhythmic motor patterns are driven by common premotor and motor neurons in a simple vertebrate spinal cord. *J. Neurosci.* **13**, 4456–4469 (1993).
 86. Earhart, G. M. & Stein, P. S. G. Step, swim and scratch motor patterns in the turtle. *J. Neurophysiol.* **84**, 2181–2190 (2000).
 87. Lieske, S. P., Thoby-Brisson, M., Telgkamp, P. & Ramirez, J. M. Reconfiguration of the neural network controlling multiple breathing patterns: eupnea, sighs and gasps. *Nature Neurosci.* **3**, 600–607 (2000).
 88. Dickinson, P. S., Mecsas, C. & Marder, E. Neuropeptide fusion of two motor-pattern generator circuits. *Nature* **344**, 155–158 (1990).
 89. Scholz, N. L., de Vente, J., Truman, J. W. & Graubard, K. Neural network partitioning by NO and cGMP. *J. Neurosci.* **21**, 1610–1618 (2001).
 90. Hemptle, C. M., Vincent, P., Adams, S. R., Tsien, R. Y. & Selverston, A. I. Spatio-temporal dynamics of cyclic AMP signals in an intact neural circuit. *Nature* **384**, 166–169 (1996).
 91. Swensen, A. M. & Marder, E. Multiple peptides converge to activate the same voltage-dependent current in a central pattern-generating circuit. *J. Neurosci.* **20**, 6752–6759 (2000).
 92. Marder, E. & Calabrese, R. L. Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
 93. Nagy, F. & Card, P. A rhythmic modulatory gating system in the stomatogastric nervous system of *Homarus gammarus*. II. Modulatory control of the pyloric CPG. *J. Neurophysiol.* **71**, 2490–2502 (1994).
 94. Perrins, R. & Weiss, K. R. Compartmentalization of information processing in an aplysia feeding circuit interneuron through membrane properties and synaptic interactions. *J. Neurosci.* **18**, 3977–3989 (1998).
 95. Wood, D. E., Stein, W. & Nusbaum, M. P. Projection neurons with shared cotransmitters elicit different motor patterns from the same neural circuit. *J. Neurosci.* **20**, 8943–8953 (2000).
 96. Wood, D. E. & Nusbaum, M. P. Extracellular peptidase activity tunes motor pattern modulation. *J. Neurosci.* **22**, 4185–4195 (2002).
 97. Richards, K. S. & Marder, E. The actions of crustacean cardioactive peptide on adult and developing stomatogastric ganglion motor patterns. *J. Neurobiol.* **44**, 31–44 (2000).
 98. Le Feuvre, Y., Fenelon, V. S. & Meyrand, P. Ontogeny of modulatory inputs to motor networks: early established projection and progressive neurotransmitter acquisition. *J. Neurosci.* **21**, 1313–1326 (2001).
 99. O'Donovan, M. J. The origin of spontaneous activity in developing networks of the vertebrate nervous system. *Curr. Opin. Neurobiol.* **9**, 94–104 (1999).
 100. Wong, R. O. L. Retinal waves and visual system development. *Annu. Rev. Neurosci.* **22**, 29–47 (1999).

Acknowledgements

We thank E. Marder for helpful discussions and R. Harris-Warrick for feedback on an earlier version of the manuscript. Research support in the our laboratory is from National Institute of Neurological Disorders and Stroke grants.

What songbirds teach us about learning

Michael S. Brainard & Allison J. Doupe

W. M. Keck Center for Integrative Neuroscience and Departments of Physiology and Psychiatry, University of California, San Francisco, California 94143, USA (e-mail: msb@phy.ucsf.edu)

Bird fanciers have known for centuries that songbirds learn their songs. This learning has striking parallels to speech acquisition: like humans, birds must hear the sounds of adults during a sensitive period, and must hear their own voice while learning to vocalize. With the discovery and investigation of discrete brain structures required for singing, songbirds are now providing insights into neural mechanisms of learning. Aided by a wealth of behavioural observations and species diversity, studies in songbirds are addressing such basic issues in neuroscience as perceptual and sensorimotor learning, developmental regulation of plasticity, and the control and function of adult neurogenesis.

The study of birdsong exemplifies a neuroethological approach to understanding brain function, in which a detailed knowledge of naturally occurring behaviours can inform and guide the search for underlying neural mechanisms. Songbirds also illustrate the neuroethological tenet that 'specialist' animals can provide revealing examples of basic processes shared by less specialized animals. For example, songbirds display complex perceptual learning, in which experience interacts with inborn predispositions to learn. They engage in sophisticated motor skill learning, guided by performance-based feedback. Their capacity for song learning is restricted to a sensitive period in development, the timing of which depends on experience as well as hormones, and varies between species. To subserve vocal learning, songbirds have evolved a discrete set of brain structures, which include specializations of widely conserved vertebrate circuitry, such as basal ganglia networks. The structures involved in song learning and production have also revealed some of the most compelling examples of adult neurogenesis and its regulation. Finally, the properties of song learning make it a model not only for general sensory and motor learning, but also for human speech learning, providing one of the few model systems for the human capacity to acquire vocal behaviour.

In this highly selective review of birdsong learning, we begin by outlining some behavioural observations. We then describe what can be inferred about nervous system function from these observations, as well as what has been learned about how the brain solves these behaviourally defined tasks. In some cases, inroads have been made into understanding how the nervous system carries out components of the behavioural repertoire; in most, however, our understanding remains incomplete. Nevertheless, songbirds provide a system where observation of naturally occurring

behaviours has delineated a series of questions of general relevance to learning, in a context where it is highly tractable to elucidate neural mechanisms.

Behavioural basis of vocal learning

The importance of hearing

The scientific study of birdsong began in the late 1950s, with Thorpe¹ and Marler². They showed that birds, taken from the wild as eggs or nestlings and tutored with songs of unrelated adults of the same species (conspecifics), ultimately produce songs that resemble the tutor songs (Fig. 1a–c). In contrast, birds raised in acoustic isolation from conspecific males produce very abnormal 'isolate' songs (Fig. 1d). Moreover, some songbirds, like humans, have learned, geographically restricted 'dialects'². These results illustrate that hearing the sounds of others during an early 'sensory learning' period (Fig. 2) is essential to normal learning.

Songbirds must also be able to hear themselves in order to learn to vocalize normally. If birds are deafened after exposure to the songs of others, but before they begin practising their vocalizations during 'sensorimotor learning' (Fig. 2), they develop highly abnormal songs that show no evidence of learning³. However, the tutor no longer needs to be heard during this rehearsal phase. These behavioural results suggest that during the sensory phase of learning, young birds form an internal representation of song to

which they are exposed — a song 'template'. Later, during the sensorimotor phase, birds use auditory feedback to compare their developing vocalizations with the template, and guide their song modification using this comparison^{3,4}.

Birdsong and human speech

Speech and song are both complex acoustical signals (Fig. 1), and numerous features of song learning show striking parallels to human speech development⁵. Most important, humans, like songbirds, depend critically on hearing both themselves and



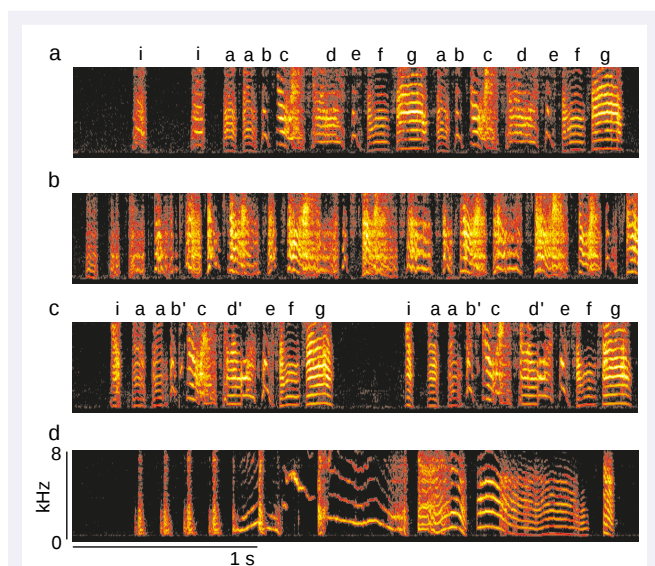


Figure 1 Birdsongs consist of ordered, often highly stereotyped strings of sounds separated by brief silent intervals. Sound energy is plotted as a function of frequency and time. Syllables are indicated by letters, and form a repeated 'motif'. **a**, Adult zebra finch song. **b**, Song of a zebra finch, tutored by the bird in **a**, at an early stage of sensorimotor learning. **c**, Song of the same bird close to song 'crystallization'. Note the similarities between this bird's song and that of its tutor. **d**, Song of a zebra finch raised in acoustic isolation. Note the overall simplicity of this song, but its general similarity of structure to other zebra finch songs.

others for normal learning. The need for experience of the sounds of other individuals is evident in the culturally transmitted languages and dialects of humans, as well as in the abnormal vocalizations of children raised without exposure to speech. The importance of auditory feedback is revealed by the profound deterioration of speech that occurs if children become deaf early or even late in childhood⁶.

The capacity for such hearing-dependent vocal learning is not widespread⁵. Apart from humans, no primates have been shown to learn their complex vocalizations. Among the rest of the mammals, only cetaceans (whales and dolphins) and some bats show evidence of vocal learning. In contrast, the vocal behaviour of the many thousands of songbird species, as well as of parrots and hummingbirds, provides a rich source of possible models for human speech learning.

Some aspects of birdsong are clearly not analogous to human speech. Although birdsong is used for communication, it does not seem to be 'language' in the sense of conveying complex meaning. What it shares with speech is the learned sensorimotor control of an elaborate vocal system. The strikingly similar requirements for this vocal learning in songbirds and humans suggest that there may be related neural mechanisms, even in brain areas that are not homologous.

Neural substrates for vocal behaviour

The behavioural studies of vocal learning indicate that there must be neural circuitry for a variety of processes, in particular (1) producing the motor commands that give rise to the complex sounds of song; (2) perceptual learning of sounds, including the memorization of the tutor song; and (3) evaluating auditory feedback relative to the internal template, and generating signals that can guide consequent modification of vocal motor output. The likely locations for these processes are a set of brain structures known as the 'song system' (Fig. 3), outlined below.

The motor pathway

Song, like speech, requires the coordinated control of vocal and respiratory musculature^{7,8}. Evidence suggests that the 'motor pathway'

of the song system (Fig. 3) generates and coordinates the patterned breathing and vocal muscle activity necessary for song production. Lesions of either of the two main nuclei of this pathway, the HVC (abbreviation used as proper name) and the robust nucleus of the archistriatum (RA), result in abnormal songs or muteness⁹, and neurophysiological activity in HVC and RA is correlated with song production^{10,11}. There is evidence for a motor hierarchy: HVC encodes higher-level song structure than does RA¹², and microstimulation in HVC causes interruption of singing and restarting of the song, whereas the same stimulation in RA disrupts only the structure of syllables without altering song patterning¹³.

Auditory areas and mechanisms

A network of forebrain auditory areas (Fig. 3) radiating from Field L, which is analogous to primary auditory cortex of mammals, is the likely source of auditory inputs to the song system^{14,15}. These high-level auditory regions may also be sites of some of the specialized operations critical to song learning. In addition, the song system itself contains some of the most complex sensory neurons known, which respond selectively to the sound of the bird's own song^{16–18} (see below).

The anterior forebrain pathway and basal ganglia function

The anterior forebrain pathway (AFP) indirectly connects the motor nuclei HVC and RA^{9,19,20} (Fig. 3). Anatomical, physiological and behavioural evidence supports the identification of this pathway as a specialized basal ganglia thalamo-'cortical' loop^{19,20}. In contrast to the song motor pathway, the AFP seems to contribute minimally to the production of stable adult song^{9,21}. However, lesions of the AFP during song learning prevent birds from developing normal adult songs^{21–23}, consistent with a function in sensory or sensorimotor learning (see below). The combination of a specialized basal ganglia circuit and a stereotyped motor output may make the AFP a particularly tractable system for revealing principles of basal ganglia function in motor learning.

Sensory learning

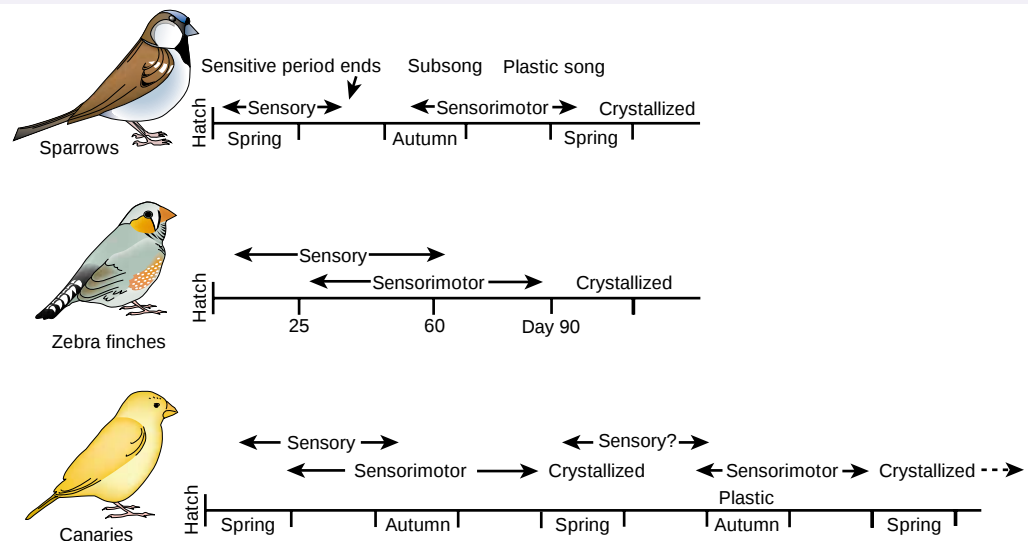
Nature versus nurture

The brain of a young songbird is not a clean slate. Without previous song exposure, young birds show greater changes in heart rate and more begging calls in response to conspecific songs than to songs of other species (heterospecific songs)^{24,25}. Moreover, although young birds are capable of copying heterospecific song, especially if it is the only model available, they will preferentially learn conspecific songs when given a choice²⁶. Finally, although birds raised in acoustic isolation sing much less complex songs than tutored birds, their songs contain some species-specific structure^{2,4} (Fig. 1d). Because most of this structure is absent in birds deafened before sensorimotor learning^{3,4}, isolate songs seem not to be pre-specified motor programs, but rather involve some sensory recognition, perhaps with respect to an 'innate template'. These behavioural studies indicate that there must be genetically determined circuitry for innate species-specific song recognition and learning.

Sensitive periods for sensory learning of song

Numerous forms of learning are subject to 'sensitive' or 'critical' periods during which experience is crucial in shaping nervous system function. One well known sensitive period is that for human speech; after early adolescence, it is difficult to learn to produce the sounds of a new language with the competence of a native speaker⁵. Many songbird species have a similar sensitive period (Fig. 2). For instance, recorded tutor songs presented to a white-crowned sparrow after 100 days of age do not appear in the bird's adult song. Similarly, white-crowned sparrows raised in isolation to 100 days of age and subsequently exposed to taped tutors still produce abnormal isolate songs as adults^{2,4}. For songbirds (and perhaps for humans as well), it seems that it is the capacity for sensory learning that declines with age (see ref. 5 for a review).

Figure 2 Timelines for song learning. **a**, In many seasonal species, such as the white-crowned sparrow, the sensory and sensorimotor phases of learning can be separated in time. The initial vocalizations, or 'subsong', produced by young birds are variable and generic across individuals, akin to the babbling of human infants. Subsong gradually evolves into 'plastic song', which remains highly variable from one rendition to the next, but also begins to incorporate some recognizable elements of tutor songs. Plastic song is progressively refined until the bird 'crystallizes' its stable adult song. **b**, Zebra finches develop rapidly, and their two phases of learning overlap extensively. **c**, 'Open learners', such as canaries, can continue or recapitulate the initial learning process as adults.



The sensory exposure required for tutor song memorization can be surprisingly short. Nightingales can almost fully reproduce tapes of 60 songs that they have heard only once a day for 20 days²⁷, and zebra finches can learn well with less than a minute of tutor song exposure per day²⁸. In this respect, sensory learning of song resembles 'imprinting', in which animals very rapidly and irreversibly learn to recognize an animal or object of critical behavioural relevance.

Closure of the sensitive period is affected by experience

The sensitive period for song learning does not have a strict age limit. Rather, experience itself is centrally involved in closing the sensitive period. For instance, songbirds tutored with only heterospecific songs can incorporate new songs from their own species at a time when birds raised with conspecifics will no longer learn^{1,29}. For some species, even more deprivation, such as raising birds in complete isolation, can result in adults that will still incorporate new song elements^{29,30}. Thus, a lack of normal experience leaves the brain open to be shaped by the appropriate input for longer than usual. In most cases, however, plasticity seems not to last indefinitely, even in the absence of experience. Presumably, circuits poised to be shaped by activity-dependent events ultimately stabilize in some state, even if driven only by spontaneous activity.

Attentional or motivational factors also influence the timing of the sensitive period. Birds will learn from live, countersinging tutors for longer than they learn from taped tutors^{31,32}. Hormonal factors may be important as well, as manipulations that delay the onset of singing and decrease testosterone levels seem to extend the sensitive period^{1,33}.

Auditory neurons shaped by song experience

We do not yet know where and how in the brain the memory of the tutor song is stored during the process of sensory learning, nor how this memory is accessed during the evaluation of auditory feedback that guides vocal practice. However, the use of behaviourally relevant auditory stimuli has revealed neurons that clearly have been shaped by the individual bird's unique auditory experience during song learning. These 'song-selective' neurons, which are found throughout the adult male song system, respond more strongly to the sound of the bird's own song (BOS), and in some cases to the tutor song, than to other equally complex auditory stimuli, such as conspecific songs or BOS played in reverse or out of order¹⁶⁻¹⁸ (Fig. 4).

Although song-selective neurons reflect the individual bird's experience, it is not clear which aspects of that experience are

responsible for generating selectivity. In principle, these neurons might be shaped by the tutor song during sensory learning and/or by feedback of BOS during sensorimotor learning. The former possibility is especially intriguing; if tutor song selectivity arises during sensory learning, then this selectivity itself may be a manifestation of the tutor song memory. Moreover, such tutor-tuned neurons could participate directly in the subsequent evaluation of auditory feedback during sensorimotor learning: as a bird practises his song, auditory feedback from those variants that more closely resemble the tutor's song would be differentially effective in activating tutor-selective neurons. Hence, the degree of activation of these neurons during vocal practice could signal the degree of success in the young bird's attempts to mimic the tutor song.

This simple scenario, in which sensory learning generates tutor-selective neurons that can then guide feedback evaluation, faces a serious challenge. Developmental studies suggest that robust song selectivity does not emerge during sensory learning, but instead arises in parallel with the bird's own motor production. Moreover, the emerging song selectivity is characterized by greater response to BOS than to the tutor song, or by similarly strong responses to both these songs^{18,34,35}. These observations are consistent with the possibility that the critical experience that shapes song selectivity is exposure to feedback of BOS. The tutor song responses observed could arise simply because of similarity between the bird's learned song and the tutor song to which it was exposed; neurons tuned to BOS would tend to respond well to the acoustically similar tutor song (Fig. 4c).

Because of this problem of acoustic similarity, and because the tutor song is only an indirect representation of what the bird has actually memorized, the relative strength of neural responses to BOS and tutor song in normal adults cannot unambiguously reveal which experiences have shaped song selectivity³⁵ (Fig. 4c). This problem has been partly addressed by studying birds that were prevented from producing a good copy of the tutor song by denervating the vocal apparatus. These birds produce very abnormal songs, without the usual acoustic similarity to tutor song. Song-selective neurons in these birds, at least in the AFP, develop sensitivity to the sound of the abnormal songs produced by the bird³⁵. This indicates that BOS shapes song-selective neurons during sensorimotor learning.

But some AFP neurons in birds with deafferented vocal organs are strongly responsive to the tutor song as well as to BOS, despite the acoustic differences between the two³⁵. Thus, some song-selective neurons seem to reflect independently both sensory and sensorimotor learning. Such joint selectivity for BOS and tutor song could be a

useful property for song learning, which involves comparing these two stimuli. There is as yet little evidence in the song system for the simpler idea of auditory neurons with strong suprathreshold selectivity to tutor song alone.

Neurons with responses to BOS playback in anaesthetized or sleeping animals do not always show these responses when birds are awake, indicating that the strength, and perhaps the nature, of auditory responses to sounds are 'gated' by the behavioural state of the bird^{36,37}. In other sensorimotor systems, for instance locomotion in mammals or flying in insects, sensory responses related to a behaviour are 'gated' by the motor activity that generates the behaviour³⁸. That is, responses are diminished unless the animal is also engaged in the behaviour. Similarly, for songbirds as for humans, auditory feedback of self is available only when the animal is actually vocalizing. Thus, anaesthesia or sleep may artificially open a gate that is normally operated by the act of singing. Ultimately, an understanding of the neural mechanisms for evaluation of auditory feedback of BOS is likely to require recording neural activity when that feedback is produced — that is, during singing.

Forebrain auditory areas and sensory responses

Complex stimulus selectivity is also found in some auditory forebrain regions that provide input to the song system^{15,39–41}. In particular, the high-level auditory areas (Fig. 3) known as the caudomedial neostriatum (NCM) and the caudal portion of the ventral hyperstriatum contain neurons that show more immediate early gene induction or neurophysiological activity in response to conspecific songs than to heterospecific songs^{39,40}. For the most part, responses within these regions, unlike those within the song system, do not seem to be restricted specifically to BOS or tutor song stimuli. Hence, these forebrain regions may contribute to a general processing of conspecific sounds. However, one recent study found that, within NCM, some auditory responses seem to reflect the individual bird's song-learning experience^{42,43}. It therefore remains possible that some of the sensory learning of song occurs within this network of auditory forebrain areas.

This conclusion seems especially plausible as many animals that are not vocal learners, including some birds, are nevertheless capable of perceptual learning. Perceptual learning, including that of tutor song, may rely on sensory processing pathways that are phylogenetically widespread. In contrast, the sensorimotor component of vocal learning, which has appeared only rarely, may have required the evolution of specialized vocal areas such as the song system.

Assessing the functional role of brain regions in sensory learning

Lesion studies are problematic for identifying brain regions that are specifically involved in the sensory phase of song learning. This is because the main assay for what a bird has memorized is the song that the bird ultimately produces; any song abnormalities arising from lesions are therefore difficult to attribute specifically to disruption of sensory learning, as opposed to disruption of subsequent sensorimotor learning or song production. One attempt to circumvent this problem in investigating the role of the AFP has been to reversibly inactivate the AFP nucleus LMAN (lateral magnocellular nucleus of the anterior neostriatum) during tutoring sessions, but not during song rehearsal⁴⁴. Song learning in these experimental birds is reduced relative to controls. However, the decrease is small, and the songs of treated birds are not isolate-like, as might be expected if song memorization were completely prevented. Nonetheless, this experiment provides perhaps the most direct evidence of involvement of a brain area in song memorization, and could be extended usefully to testing the role of other brain areas.

Another approach is to study the effects of lesions in purely perceptual tasks, where aspects of sensory learning can be measured independently of song production. Although such studies have not addressed the issue of memorization of tutor song, they have found that lesions of song nuclei, including HVC and LMAN,

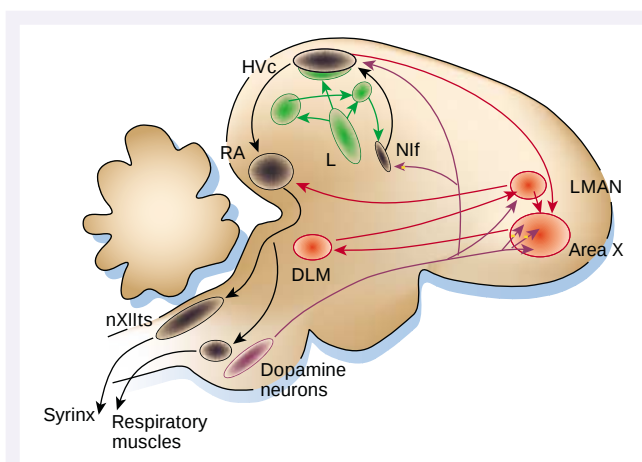


Figure 3 Neural substrates for learning: the song system. The motor pathway (black) is necessary for normal song production throughout life, and includes HVC (abbreviation used as proper name) and the robust nucleus of the archistriatum (RA)⁹. RA projects to the tracheosyringeal portion of the hypoglossal nucleus (nXlts), which controls the bird's vocal organ or syrinx, and to nuclei involved in control of respiration during song^{7–9}. Additional nuclei afferent to HVC, including the nucleus interfacialis (Nif), are likely to be part of the motor pathway, but their role is less clear. HVC sends a second projection to the anterior forebrain pathway (AFP, red). The AFP includes Area X, which is homologous to mammalian basal ganglia^{19,20}, the medial nucleus of the dorsolateral thalamus (DLM), and the lateral magnocellular nucleus of the anterior neostriatum (LMAN; a frontal cortex-like nucleus). LMAN sends a projection back into the motor pathway at the level of RA. Like basal ganglia in other vertebrates, Area X is the target of strong midbrain dopamine projections¹⁹; LMAN, HVC and Nif also receive dopamine inputs (purple). The Field L complex is the avian primary forebrain auditory area and projects to a complex network of higher auditory areas¹⁴ (green), including the caudomedial neostriatum and caudal portion of the ventral hyperstriatum (not labelled). Auditory inputs likely enter the song system at the level of Nif and possibly HVC¹⁵.

interfere with the performance of birds in tasks that require song memorization and discrimination^{45–47}.

Cellular and synaptic changes correlated with sensory learning

Tutor song memorization could be distributed across a number of brain areas, but because disruptions of the AFP affect learning, many studies have focused on this circuit in the search for neural mechanisms underlying the sensitive period for sensory learning⁴⁸. In zebra finches, the AFP and its connection to RA undergo numerous regressive changes by 60 days of age, when the sensitive period closes in this species. The synapses from LMAN to the motor pathway decrease in number when HVC innervates RA⁴⁹, and the initially coarse topographic projection from LMAN to RA undergoes refinement⁵⁰. Elimination of connections is also prominent within the AFP itself: LMAN neuron spine density decreases between 25 and 60 days of age⁵¹, and thalamic arbors in LMAN are pruned⁵². This is accompanied by decreased *N*-methyl-D-aspartate (NMDA) receptors in LMAN⁵³, faster NMDA currents at synapses from thalamus to LMAN³⁰, and loss of activity-dependent synaptic potentiation and depression at synapses within LMAN⁵⁴.

These regressive changes could potentially underlie an experience-dependent narrowing of song responsiveness as birds encode a particular tutor song memory. But in zebra finches, the period of sensory learning also overlaps with the onset of vigorous singing, sensorimotor rehearsal and refinement of auditory selectivity for BOS (Fig. 2), making it difficult to specifically attribute any changes to sensory learning. In addition, the song system is still developing during this time, such that many of the observed changes could reflect developmental events that are independent of learning. So far, only a small number of observations have been tested and found to

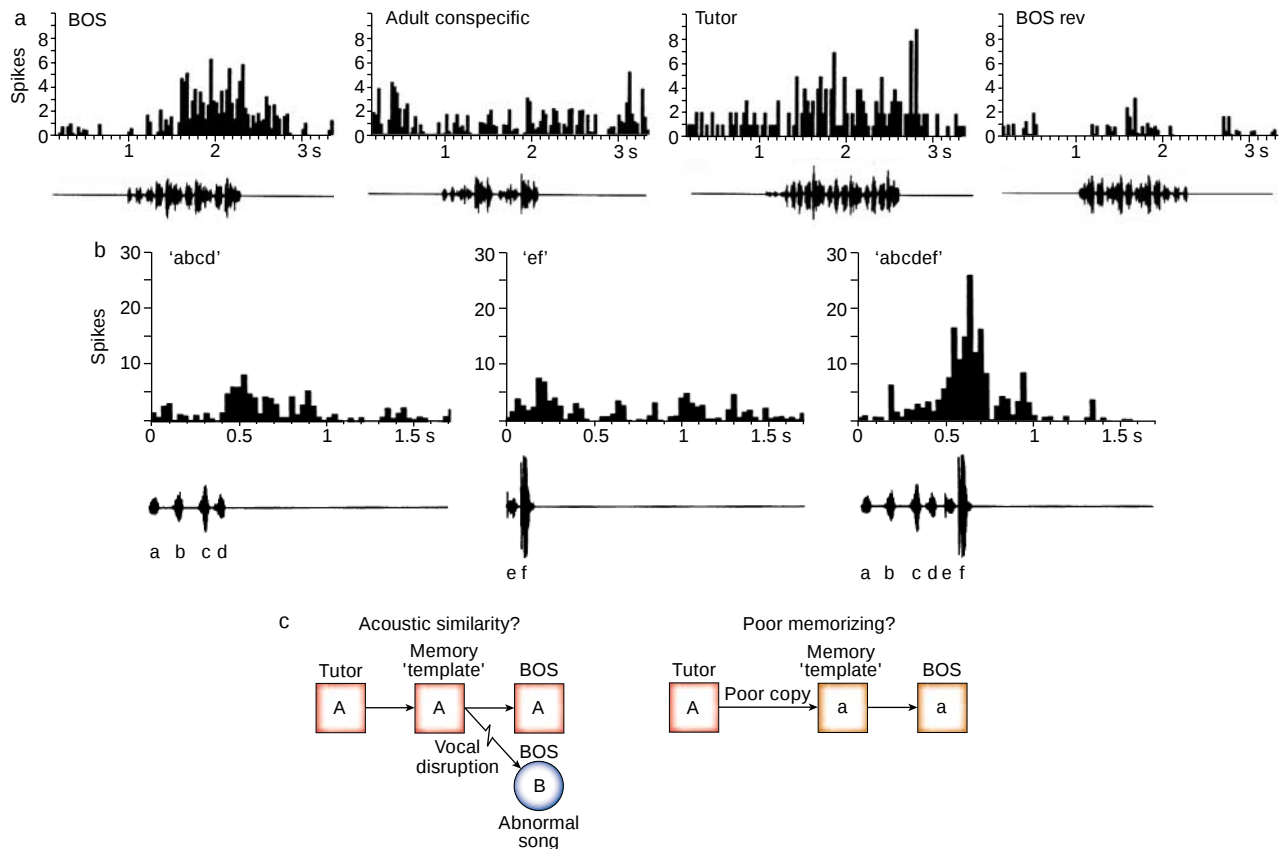


Figure 4 Song selectivity. **a**, Song-selective neurons respond better to the bird's own song (BOS), and in some cases the tutor song, than to equally complex conspecific songs. They also respond more strongly to the song in the forward order than to the same song reversed (BOS rev). **b**, Song-selective neurons are highly nonlinear, and in some cases 'combination sensitive'^{16–18}, responding better to a combination of sounds than to any of the sounds in isolation. Intracellular recordings have begun to address the cellular and synaptic mechanisms that must underlie such complex spectrally and temporally tuned responses to song^{99,100}, and raise the intriguing possibility that there are different representations of BOS in different classes of song-selective neurons¹⁰⁰.

c, If the bird both memorizes and reproduces the tutor song 'A' very faithfully, as in the left panel, the tutor song will be virtually identical to BOS. Neuronal responsiveness to these two songs will not distinguish which experience was primary in shaping the neurons. If, however, the bird memorizes poorly ('a') but copies the inaccurate template faithfully into its own song, as in the right panel, BOS will be the best reflection of what the bird actually stored in memory. Neuronal selectivity for BOS in this case will reflect what was memorized in response to tutor song exposure. Artificially altering the bird's song production (creating abnormal song 'B'³⁵) can eliminate similarity between BOS and tutor song, and help identify which of these stimuli shape the neural response.

correlate with learning rather than developmental age. For example, the elimination of spines normally seen in LMAN of zebra finches by day 60 does not occur in birds raised without tutors⁵¹, implying that spine loss in LMAN may be a cellular consequence of sensory experience and learning. In contrast, although isolation rearing enables late learning, it delays, but does not prevent, shortening of NMDA-receptor kinetics at thalamus–LMAN synapses³⁰. The ability of isolates to learn new songs clearly indicates that changes in NMDA-receptor kinetics, at least at thalamus–LMAN synapses, do not prevent song learning in the way that closure of the sensitive period does.

Sensorimotor learning

During sensorimotor song learning (Fig. 2), motor circuitry is gradually shaped by performance-based feedback to produce an adaptively modified behaviour. This feedback is critical throughout motor learning—at any point prior to crystallization, elimination of auditory feedback by deafening not only arrests the progression of sensorimotor learning, but also can lead to a rapid deterioration of song, including the loss of previously learned elements⁵⁵.

The AFP and sensorimotor learning

Because AFP lesions, like deafening, dramatically disrupt song development, this circuit may function in sensorimotor learning of

song. Studies of the AFP in adult birds provide several suggestions about what such a function might be. AFP lesions prevent a variety of changes to adult song that can otherwise be driven by manipulations of experience such as deafening or unilateral denervation of the vocal musculature^{56,57}. Because these manipulations create feedback of song that differs from what is expected, the effects of AFP lesions are consistent with the hypothesis that the AFP functions throughout life to evaluate auditory feedback of song with respect to the desired output, and to instruct changes in the motor pathway. Alternatively, or in addition, the AFP may act more permissively in motor pathway plasticity, enabling change without providing specific guidance. This is consistent with the known trophic role of the AFP in the survival, growth and innervation of RA neurons^{58,59}. Regardless of mechanism, the AFP seems to be crucial in both song learning and plasticity of adult song.

In adult birds, the AFP is active during singing^{60,61}, with premotor activity resembling that in the motor nucleus HVC. This suggests that AFP activity originates from HVC, and represents in part an 'efference copy' of the premotor signals sent to the motor output pathway. Efference copies of motor activity are common in sensorimotor systems, and may generate predictions about the expected sensory consequences of motor commands. Such an efference copy

of song motor activity in the AFP could be particularly useful during sensorimotor learning^{62,63}.

Song-selective neurons as motor neurons

Song-selective neurons not only respond to complex sensory signals, but also can be active during motor production^{61,63}. For instance, the same RA neurons that exhibit song-selective responses in sleeping birds are active during singing⁶³. There is a remarkable correspondence between these neurons' auditory responses to song and their premotor activity — playback of one set of syllables triggers an auditory response that resembles the premotor activity for the next syllable in the song. Thus the auditory response can be considered a prediction of the motor command for the following syllable. These results raise the possibility that song-selective neurons in both the motor pathway and the AFP are critically involved in linking sensory and motor representations in the song system^{62,63}.

Crystallization and selection of song

Studies of developing or 'plastic' song reveal that some bird species sing more sounds as juveniles than are ultimately preserved in their adult song. For example, birds exposed to multiple tutor songs may sing plastic songs that include virtually complete renditions of each tutor song, even though they will eventually sing only one of these⁶⁴. During sensorimotor learning, birds are thus not only learning how to produce previously memorized sounds, but are also selecting which of these acquired skills will ultimately be expressed. It is clear that external stimuli, including social interactions, are important in this process. For example, white-crowned sparrows singing multiple juvenile songs in the field crystallize the one that is most similar to songs of other birds in the vicinity, and this selection can be reproduced in the laboratory using song playback⁶⁴. Social influences on song crystallization are not only acoustic; male cowbirds will preferentially retain songs that prove effective in eliciting courtship displays from females⁶⁵.

How messages about song quality are conveyed to the song motor pathway is unknown. However, in adult birds, social context strongly modulates the level of activity within the AFP of the song system during song production^{66,67}. Similar modulation during the final stages of sensorimotor learning might have a role in song selection. Because reinforcement signals provided by social interactions can be non-acoustic, ascending dopaminergic projections from the midbrain ventral tegmental area could be involved (these are thought to have a highly conserved role in mediating effects of reward and reinforcement). In songbirds, dopaminergic pathways send a particularly dense projection to nuclei of the song system, especially Area X of the AFP^{19,20}, and are thus well situated to modulate song learning.

Regulation of sensorimotor plasticity

Although there is strong evidence for sensitive periods for the sensory phase of song learning, it is less straightforward to determine whether there is similar regulation of sensorimotor learning. The stability of adult song in many species raises the possibility that motor circuitry becomes 'crystallized' and unchangeable, although this could also reflect continued matching of song output to an unchanging sensory template. Pytte and Suthers⁶⁸ showed that transiently disrupting motor production with botulinum toxin in very young or adult birds has no lasting effects on song. However, if birds are prevented from vocalizing normally during the later stages of sensorimotor learning, just before or during song crystallization, song becomes permanently distorted⁶⁸. This pre-crystallization period may thus represent a motor sensitive period during which it is crucial that birds have normal opportunity for rehearsal.

More pronounced perturbations of the motor periphery make it clear that adult song production does not become completely refractory to the influence of experience. Crushing the nerve that innervates the syrinx⁶⁹, or interfering (reversibly) with the mechanics of syringeal movement⁷⁰ not only acutely disrupts song, but, unlike botulinum toxin, eventually leads to permanent elimination of song elements

and gross changes to the temporal pattern of song. Such changes must reflect alterations in the central motor pathway of adults.

It is not certain whether these disruptions of motor production are caused by damage to the motor periphery *per se*, or by the resulting abnormal auditory feedback (see ref. 71 for full discussion). Manipulations of auditory feedback that can be imposed without interfering with motor production, such as deafening or reversible disruption of auditory feedback, can also lead to deterioration of adult song^{55,72–75}. This indicates that there is not a complete loss of plasticity in the motor pathway even in response to a primarily sensory manipulation, and suggests that auditory feedback continues to exert an important influence on adult song.

The degree of song deterioration after hearing loss is much less severe in adulthood than in juveniles^{55,72}, and the effects of deafening continue to wane even after the apparent crystallization of adult song. For zebra finches, the consequences of deafening for adult song are much greater shortly after song crystallization than when birds are deafened progressively later over the ensuing months^{76,77}. Humans show a similar dependence on auditory feedback. In adults, speech gradually deteriorates after hearing loss, but exhibits progressively less deterioration as hearing loss occurs later over the second, third and fourth decades of life⁶. These findings suggest that for both birds and humans, even after sensorimotor learning seems to be complete, there is nevertheless a continuing, covert stabilization of adult vocalizations.

Sleep may be important in such consolidation of song, or in song learning more generally. During sleep, some of the spontaneous bursting of adult RA neurons is similar in its pattern to the activity of the same neurons during singing, and thus perhaps reflects 'replay' of activity that occurred during the day⁶³. Such replay could be involved in 'off-line' alteration or strengthening of connections in the neural network for song; this would be consistent with reports hypothesizing a role for sleep in learning and consolidation of memory⁷⁸.

Hormones and sensorimotor plasticity

Songbirds vary widely in the degree to which they can modify their song in adulthood. 'Closed-learners' like the zebra finch or white-crowned sparrow pass through a single period of song learning and then normally retain an essentially unchanging song throughout life. In contrast, 'open-learners', like canaries or starlings, initially pass through sensory and sensorimotor learning resulting in stable adult song, but then can continue to learn^{79,80} (Fig. 2). In canaries this occurs seasonally; after a winter period of song variability, they produce a stable song each spring into which new elements have been incorporated⁷⁹. Comparisons across species that have different capacities for learning, and within species at times when learning is differentially enabled, have the potential to reveal what factors regulate nervous system plasticity.

Steroid hormones may be one such factor. Sex steroids are well known to shape the development of the sexually dimorphic song system^{4,81,82}, but they also seem to influence learning more directly¹⁹. Testosterone levels rise during sensorimotor learning, in parallel with song crystallization; they are high in springtime, when song is stable, and low in late summer and autumn⁷⁹. Testosterone treatment can precipitate premature crystallization of abnormally simple song^{83,84}. Conversely, depletion of testosterone can delay or prevent crystallization of song^{85,86}. Testosterone causes numerous structural and electrophysiological changes in song system neurons^{19,87}, which could influence plasticity. Because steroid hormone receptors are particularly enriched in the song motor pathway and LMAN^{19,87}, testosterone could be acting directly at these sites. Alternatively, because androgens increase singing, the effects of testosterone on the song system could reflect the indirect consequences of increased motor performance⁸⁸.

New neurons in adult brains

The song system has contributed greatly to our understanding of neuron generation in adulthood, again aided both by behavioural

knowledge and by the discrete circuit underlying song. The initial suggestion that new neurons were born in the adult mammalian brain⁸⁹ met with resistance. But in the early 1980s, while searching for possible mechanisms underlying seasonal changes in song and in the volume of the song control nucleus HVC, Goldman and Nottebohm⁹⁰ discovered striking amounts of adult neurogenesis in the songbird forebrain, including HVC. Because many new neurons were added to a well-defined circuit, Nottebohm and colleagues were able to provide compelling evidence for neurogenesis, including electron microscopy, retrograde neuronal labelling, and neurophysiological recordings from newly generated neurons^{91,92}. Only recently has it become more generally accepted that adult neurogenesis also occurs in mammalian brains⁹³.

Neurogenesis is regulated in adult songbirds (as it seems to be in mammals): not all brain areas receive new neurons, and not all neuronal types are readily generated anew in adulthood⁹². Although neurogenesis occurs in non-song learners and in many areas of the avian forebrain, including the hippocampus, studies within the well-delineated song system have facilitated recognition of the regulation of neurogenesis. For instance, HVC has two intermingled populations of long-range projection neurons, one projecting to the motor nucleus RA, and the other targeting the basal ganglia nucleus Area X (Fig.3). Only RA-projecting neurons are born in adulthood⁹², and selective killing of these neurons in adult birds results in a peak of increased insertion of new neurons into HVC⁹⁴. In contrast, killing adult Area X-projecting neurons does not result in more neurogenesis. Newly generated neurons are produced when Area X-projecting neurons are killed in young birds, but the new neurons are all RA-projecting⁹⁴. Thus the birth or recruitment of neurons is sensitive to injury or vacancy signals from the brain, but not all neuron types can be generated with equal ease.

Many more neurons are born in adulthood than ultimately survive. However, studies in adult songbirds have also led to some of the best evidence for prolonged survival of new neurons, with newly generated RA-projecting neurons in HVC living for at least 8 months⁹¹. In addition, these neurons send long-range projections to their appropriate targets within pre-existing, myelinated adult circuits⁹¹. Understanding the molecular signals that allow this growth and targeting should inform the search for factors important for the migration and connection of neurons in adult mammalian brain and spinal cord.

In songbirds, the evidence that neurogenesis and neuronal recruitment is sensitive to experience and environmental cues is multifaceted and compelling. Recruitment of new neurons is clearly seasonally and hormonally regulated. There are large differences in the number of new neurons observed after injection of [³H]thymidine in the autumn compared with the spring, and treatment of birds with testosterone dramatically increases the insertion and/or survival of new neurons in HVC, without affecting the rate of neurogenesis in the ventricular zone^{95,96}. Eliminating auditory input by deafening animals also alters the number of new neurons in HVC⁹⁷, as does the act of singing⁸⁸.

Perhaps the most intriguing question regarding adult neurogenesis is whether newly generated neurons have an important function, especially in learning. In songbirds, the strong association between a stereotyped learned vocal behaviour and discrete brain areas should make it feasible to test whether there is a causal link between new neurons and song plasticity. Interestingly, there are peaks of neuron loss and replacement in canary HVC that correlate with seasonal periods of song instability and restabilization, respectively⁹². However, new neurons are also inserted in other seasonal species that do not change their songs⁹⁸. Thus, critical experiments remain to be done, to see whether eliminating neurogenesis prevents song learning or modification.

The observation that many birds that incorporate new neurons nonetheless have an unchanging adult song is of interest in its own right. Even the disruption of adult zebra finch song that is triggered

by ablation of RA-projecting neurons (and is followed by increased neuronal recruitment) is succeeded by gradual recovery of song to its pre-ablation state⁹⁴. Songbirds thus provide an example of a system where learned capacities and memories persist despite neural turnover. The mechanisms underlying such resilience to cell loss and replacement may be particularly relevant in the future as we attempt to repair the adult brain by inserting new neurons.

Future directions

This review highlights the richness of song learning behaviour, and the many questions of general relevance to learning and memory posed by behavioural studies. Much about the neural foundations of this behaviour remains unexplored, and we hope to have conveyed here the potential waiting to be tapped.

Songbirds have many experimental advantages, including their small size, relatively rapid development, and the ease both of altering their experience and of recording brain activity during behaviour. These features, together with the specialized brain areas for song, give this system the potential to elucidate neural mechanisms of learning from the systems down to the cellular and molecular levels.

It is a disadvantage, for molecular studies in particular, that birds do not possess the genetic tractability of animals such as mice and flies. However, the naturally occurring differences in learning between species effectively provide opportunities to study mechanisms of phenotypic variation. Moreover, new tools for 'non-genetic' animals are under development, and songbirds may prove to be relatively easy vertebrates to manipulate. Indeed, as additional organisms are selected for genome sequencing, we would be well served by targeting animals such as songbirds. Ultimately, if we are to address questions of complex natural behaviour including our own, we must tackle, at all levels of analysis, systems where evolution has resulted in elaborate, learned behavioural capacities and the corresponding neural mechanisms. □

1. Thorpe, W. H. *Bird-Song* (Cambridge Univ. Press, Cambridge, 1961).
2. Marler, P. A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* **71**, 1–25 (1970).
3. Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
4. Konishi, M. Birdsong: from behavior to neuron. *Annu. Rev. Neurosci.* **8**, 125–170 (1985).
5. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
6. Cowie, R. & Douglas-Cowie, E. *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Mouton de Gruyter, Berlin, 1992).
7. Wild, J. M. Neural pathways for the control of birdsong production. *J. Neurobiol.* **33**, 653–670 (1997).
8. Suthers, R. A., Goller, F. & Pytte, C. The neuromuscular control of birdsong. *Phil. Trans. R. Soc. Lond. B* **354**, 927–939 (1999).
9. Nottebohm, F., Stokes, T. M. & Leonard, C. M. Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* **165**, 457–486 (1976).
10. McCasland, J. S. Neuronal control of bird song production. *J. Neurosci.* **7**, 23–39 (1987).
11. Chi, Z. & Margoliash, D. Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* **32**, 899–910 (2001).
12. Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
13. Vu, E. T., Mazurek, M. E. & Kuo, Y. C. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14**, 6924–6934 (1994).
14. Vates, G. E., Broome, B. M., Mello, C. V. & Nottebohm, F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taeniopygia guttata*). *J. Comp. Neurol.* **366**, 613–642 (1996).
15. Janata, P. & Margoliash, D. Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *J. Neurosci.* **19**, 5108–5118 (1999).
16. Margoliash, D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* **3**, 1039–1057 (1983).
17. Margoliash, D. & Fortune, E. S. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J. Neurosci.* **12**, 4309–4326 (1992).
18. Doupe, A. J. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.* **17**, 1147–1167 (1997).
19. Bottjer, S. W. & Johnson, F. Circuits, hormones, and learning: vocal behavior in songbirds. *J. Neurobiol.* **33**, 602–618 (1997).
20. Luo, M., Ding, L. & Perkel, D. J. An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *J. Neurosci.* **21**, 6836–6845 (2001).
21. Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901–903 (1984).
22. Sohrabji, F., Nordeen, E. J. & Nordeen, K. W. Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behav. Neural Biol.* **53**, 51–63 (1990).
23. Scharff, C. & Nottebohm, F. A comparative study of the behavioral deficits following lesions of

- various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896–2913 (1991).
24. Dooling, R. & Searcy, M. Early perceptual selectivity in the swamp sparrow. *Dev. Psychobiol.* **13**, 499–506 (1980).
 25. Nelson, D. A. & Marler, P. Innate recognition of song in white-crowned sparrows: a role in selective vocal learning? *Anim. Behav.* **46**, 806–808 (1993).
 26. Marler, P. & Peters, S. The role of song phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. *Ethology* **77**, 125–149 (1988).
 27. Hultsch, H. & Todt, D. Song acquisition and acquisition constraints in the nightingale, *Luscinia megarhynchos*. *Naturwissenschaften* **76**, 83–85 (1989).
 28. Tchernichovski, O., Lints, T., Mitra, P. P. & Nottebohm, F. Vocal imitation in zebra finches is inversely related to model abundance. *Proc. Natl Acad. Sci. USA* **96**, 12901–12904 (1999).
 29. Slater, P. J. B., Eales, L. A. & Clayton, N. S. Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. *Adv. Study Behav.* **18**, 1–34 (1988).
 30. Livingston, F. S., White, S. A. & Mooney, R. Slow NMDA-EPSCs at synapses critical for song development are not required for song learning in zebra finches. *Nature Neurosci.* **3**, 482–488 (2000).
 31. Baptista, L. F. & Petrinovich, L. Song development in the white-crowned sparrow: social factors and sex differences. *Anim. Behav.* **34**, 1359–1371 (1986).
 32. Nordby, J. C., Campbell, S. E., Burt, J. M. & Beecher, M. D. Social influences during song development in the song sparrow: a laboratory experiment simulating field conditions. *Anim. Behav.* **59**, 1187–1197 (2000).
 33. Whaling, C. S., Soha, J. A., Nelson, D. A., Lasley, B. & Marler, P. Photoperiod and tutor access affect the process of vocal learning. *Anim. Behav.* **56**, 1075–1082 (1998).
 34. Volman, S. F. Development of neural selectivity for birdsong during vocal learning. *J. Neurosci.* **13**, 4737–4747 (1993).
 35. Solis, M. M. & Doupe, A. J. Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. *J. Neurosci.* **19**, 4559–4584 (1999).
 36. Schmidt, M. F. & Konishi, M. Gating of auditory responses in the vocal control system of awake songbirds. *Nature Neurosci.* **1**, 513–518 (1998).
 37. Dave, A. S., Yu, A. C. & Margoliash, D. Behavioral state modulation of auditory activity in a vocal motor system. *Science* **282**, 2250–2254 (1998).
 38. Pearson, K. G. Common principles of motor control in vertebrates and invertebrates. *Annu. Rev. Neurosci.* **16**, 265–297 (1993).
 39. Mello, C. V. & Clayton, D. F. Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J. Neurosci.* **14**, 6652–6666 (1994).
 40. Stripling, R., Kruse, A. A. & Clayton, D. F. Development of song responses in the zebra finch caudomedial neostriatum: role of genomic and electrophysiological activities. *J. Neurobiol.* **48**, 163–180 (2001).
 41. Theunissen, F. E., Sen, K. & Doupe, A. J. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* **20**, 2315–2331 (2000).
 42. Bolhuis, J. J., Zijlstra, G. G., den Boer-Visser, A. M. & Van Der Zee, E. A. Localized neuronal activation in the zebra finch brain is related to the strength of song learning. *Proc. Natl Acad. Sci. USA* **97**, 2282–2285 (2000).
 43. Marler, P. & Doupe, A. J. Singing in the brain. *Proc. Natl Acad. Sci. USA* **97**, 2965–2967 (2000).
 44. Basham, M. E., Nordeen, E. J. & Nordeen, K. W. Blockade of NMDA receptors in the anterior forebrain impairs sensory acquisition in the zebra finch. *Neurobiol. Learn. Mem.* **66**, 295–304 (1996).
 45. Scharff, C., Nottebohm, F. & Cynx, J. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J. Neurobiol.* **36**, 81–90 (1998).
 46. Burt, J. M., Lent, K. L., Beecher, M. D. & Brenowitz, E. A. Lesions of the anterior forebrain song control pathway in female canaries affect song perception in an operant task. *J. Neurobiol.* **42**, 1–13 (2000).
 47. Gentner, T. Q., Hulse, S. H., Bentley, G. E. & Ball, G. F. Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.* **42**, 117–133 (2000).
 48. Nordeen, K. W. & Nordeen, E. J. Anatomical and synaptic substrates for avian song learning. *J. Neurobiol.* **33**, 532–548 (1997).
 49. Herrmann, K. & Arnold, A. P. The development of afferent projections to the robust archistriatal nucleus in male zebra finches: a quantitative electron microscopic study. *J. Neurosci.* **11**, 2063–2074 (1991).
 50. Iyengar, S., Viswanathan, S. S. & Bottjer, S. W. Development of topography within song control circuitry of zebra finches during the sensitive period for song learning. *J. Neurosci.* **19**, 6037–6057 (1999).
 51. Wallhauser-Franke, E., Nixdorf-Bergweiler, B. E. & DeVoogd, T. J. Song isolation is associated with maintaining high spine frequencies on zebra finch LMAN neurons. *Neurobiol. Learn. Mem.* **64**, 25–35 (1995).
 52. Iyengar, S. & Bottjer, S. Development of individual axon arbors in a thalamocortical circuit necessary for song learning in zebra finches. *J. Neurosci.* **22**, 901–911 (2002).
 53. Aamodt, S. M., Kozlowski, M. R., Nordeen, E. J. & Nordeen, K. W. Distribution and developmental change in [3H]MK-801 binding within zebra finch song nuclei. *J. Neurobiol.* **23**, 997–1005 (1992).
 54. Boettiger, C. A. & Doupe, A. J. Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* **31**, 809–818 (2001).
 55. Price, P. H. Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol.* **93**, 268–277 (1979).
 56. Williams, H. & Mehta, N. Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *J. Neurobiol.* **39**, 14–28 (1999).
 57. Brainard, M. S. & Doupe, A. J. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762–766 (2000).
 58. Akutagawa, E. & Konishi, M. Two separate areas of the brain differentially guide the development of a song control nucleus in the zebra finch. *Proc. Natl Acad. Sci. USA* **91**, 12413–12417 (1994).
 59. Kittelberger, J. M. & Mooney, R. Lesions of an avian forebrain nucleus that disrupt song development alter synaptic connectivity and transmission in the vocal premotor pathway. *J. Neurosci.* **19**, 9385–9398 (1999).
 60. Jarvis, E. D. & Nottebohm, F. Motor-driven gene expression. *Proc. Natl Acad. Sci. USA* **94**, 4097–4102 (1997).
 61. Hessler, N. A. & Doupe, A. J. Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches. *J. Neurosci.* **19**, 10461–10481 (1999).
 62. Troyer, T. & Doupe, A. J. An associational model of birdsong sensorimotor learning. I. Efference copy and the learning of song syllables. *J. Neurophysiol.* **84**, 1204–1223 (2000).
 63. Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290**, 812–816 (2000).
 64. Nelson, D. A. & Marler, P. Selection-based learning in bird song development. *Proc. Natl Acad. Sci. USA* **91**, 10498–10501 (1994).
 65. West, M. J. & King, A. P. Female visual displays affect the development of male song in the cowbird. *Nature* **334**, 244–246 (1988).
 66. Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A. & Nottebohm, F. For whom the bird sings: context-dependent gene expression. *Neuron* **21**, 775–788 (1998).
 67. Hessler, N. A. & Doupe, A. J. Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neurosci.* **2**, 209–211 (1999).
 68. Pytte, C. L. & Suthers, R. A. Sensitive period for sensorimotor integration during vocal motor learning. *J. Neurobiol.* **42**, 172–189 (2000).
 69. Williams, H. & McKibben, J. R. Changes in stereotyped central motor patterns controlling vocalization are induced by peripheral nerve injury. *Behav. Neural Biol.* **57**, 67–78 (1992).
 70. Hough, G. & Volman, S. Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. *J. Neurosci.* **22**, 1177–1186 (2002).
 71. Brainard, M. S. & Doupe, A. J. Auditory feedback in learning and maintenance of vocal behavior. *Nature Rev. Neurosci.* **1**, 31–40 (2000).
 72. Nordeen, K. W. & Nordeen, E. J. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57**, 58–66 (1992).
 73. Leonardo, A. & Konishi, M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399**, 466–470 (1999).
 74. Okanoya, K. & Yamaguchi, A. Adult Bengalese finches (*Lonchura striata* var. *domestica*) require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.* **33**, 343–356 (1997).
 75. Woolley, S. M. & Rubel, E. W. Bengalese finches *Lonchura striata domestica* depend upon auditory feedback for the maintenance of adult song. *J. Neurosci.* **17**, 6380–6390 (1997).
 76. Lombardino, A. J. & Nottebohm, F. Age at deafening affects the stability of learned song in adult male zebra finches. *J. Neurosci.* **20**, 5054–5064 (2000).
 77. Brainard, M. S. & Doupe, A. J. Post-learning consolidation of the motor program for birdsong: stabilizing effects of age and anterior forebrain lesions. *J. Neurosci.* **21**, 2501–2517 (2001).
 78. Stickgold, R., Hobson, J. A., Fosse, R. & Fosse, M. Sleep, learning, and dreams: off-line memory reprocessing. *Science* **294**, 1052–1057 (2001).
 79. Nottebohm, F., Nottebohm, M. E., Crane, L. & Wingfield, J. C. Seasonal change in gonadal hormone levels of adult male canaries and their relation to song. *Behav. Neural Biol.* **47**, 197–211 (1987).
 80. Bohner, J., Chaiken, M. L., Ball, G. F. & Marler, P. Song acquisition in photosensitive and photorefractory male European starlings. *Horm. Behav.* **24**, 582–594 (1990).
 81. Arnold, A. P., Bottjer, S. W., Brenowitz, E. A., Nordeen, E. J. & Nordeen, K. W. Sexual dimorphisms in the neural vocal control system in song birds: ontogeny and phylogeny. *Brain Behav. Evol.* **28**, 22–31 (1986).
 82. Holloway, C. C. & Clayton, D. F. Estrogen synthesis in the male brain triggers development of the avian song control pathway *in vitro*. *Nature Neurosci.* **4**, 170–175 (2001).
 83. Korsia, S. & Bottjer, S. W. Chronic testosterone treatment impairs vocal learning in male zebra finches during a restricted period of development. *J. Neurosci.* **11**, 2362–2371 (1991).
 84. Whaling, C. S., Nelson, D. A. & Marler, P. Testosterone-induced shortening of the storage phase of song development in birds interferes with vocal learning. *Dev. Psychobiol.* **28**, 367–376 (1995).
 85. Marler, P., Peters, S., Ball, G. F., Dufty, A. M. Jr & Wingfield, J. C. The role of sex steroids in the acquisition and production of birdsong. *Nature* **336**, 770–772 (1988).
 86. Bottjer, S. W. & Hewer, S. J. Castration and antisteroid treatment impair vocal learning in male zebra finches. *J. Neurobiol.* **23**, 337–353 (1992).
 87. Ball, G. F. in *The Design of Animal Communication* (eds Hauser, M. D. & Konishi, M.) 215–253 (MIT Press, Cambridge, MA, 1999).
 88. Li, X.-C., Jarvis, E. D., Alvarez-Borda, B., Lim, D. A. & Nottebohm, F. A relationship between behavior, neurotrophin expression, and new neuron survival. *Proc. Natl Acad. Sci. USA* **97**, 8584–8589 (2000).
 89. Altman, J. & Das, G. D. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *J. Comp. Neurol.* **124**, 319–336 (1965).
 90. Goldman, S. A. & Nottebohm, F. Neuronal production, migration, and differentiation in a vocal control nucleus in the adult female canary brain. *Proc. Natl Acad. Sci. USA* **80**, 2390–2394 (1983).
 91. Kirn, J. R., Alvarez-Buylla, A. & Nottebohm, F. Production and survival of projection neurons in a forebrain vocal center of adult male canaries. *J. Neurosci.* **11**, 1756–1762 (1991).
 92. Alvarez-Buylla, A. & Kirn, J. R. Birth, migration, incorporation, and death of vocal control neurons in adult songbirds. *J. Neurobiol.* **33**, 585–601 (1997).
 93. Gould, E. & Gross, C. G. Neurogenesis in adult mammals: some progress and problems. *J. Neurosci.* **22**, 619–623 (2002).
 94. Scharff, C., Kirn, J. R., Grossman, M., Macklis, J. D. & Nottebohm, F. Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* **25**, 481–492 (2000).
 95. Rasika, S., Alvarez-Buylla, A. & Nottebohm, F. BDNF mediates the effects of testosterone on the survival of new neurons in an adult brain. *Neuron* **22**, 53–62 (1999).
 96. Brown, S. D., Johnson, F. & Bottjer, S. W. Neurogenesis in adult canary telencephalon is independent of gonadal hormone levels. *J. Neurosci.* **13**, 2024–2032 (1993).
 97. Wang, N., Aviram, R. & Kirn, J. R. Deafening alters neuron turnover within the telencephalic motor pathway for song control in adult zebra finches. *J. Neurosci.* **19**, 10554–10561 (1999).
 98. Tramontin, A. D. & Brenowitz, E. A. A field study of seasonal neuronal incorporation into the song control system of a songbird that lacks adult song learning. *J. Neurobiol.* **40**, 316–326 (1999).
 99. Lewicki, M. & Konishi, M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl Acad. Sci. USA* **92**, 5582–5586 (1995).
 100. Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J. Neurosci.* **20**, 5420–5436 (2000).

Acknowledgements

We are grateful to M. Kao for thoughtful comments on the manuscript and help with figures. Work in the authors' laboratories is supported by the NIH, the Sandler Family Supporting Foundation, NARSAD and the Grable Foundation, and the Kirsch Foundation.

Robots in invertebrate neuroscience

Barbara Webb

Centre for Cognitive and Computational Neuroscience, Department of Psychology, University of Stirling, UK (e-mail: b.h.webb@stir.ac.uk)

Can we now build artificial animals? A combination of robot technology and neuroethological knowledge is enabling the development of realistic physical models of biological systems. And such systems are not only of interest to engineers. By exploring identified neural control circuits in the appropriate functional and environmental context, new insights are also provided to biologists.

If we understand how an animal controls its behaviour, and comparable technology is available, it should be possible to build a robot that behaves the same way. Recent advances in both knowledge and technology have begun to make this possibility a realistic aim in invertebrate neuroscience. We have computing power of similar capacity to the nervous systems of insects, and can reproduce at least some of the capabilities of their sensor and actuator systems. Investigation of the behaviour and neurophysiology of invertebrates has produced sensorimotor 'circuit diagrams' to copy. There is now a growing number of studies in which hypotheses for the behavioural function of neural circuits are tested by implementing them as controllers for robots and evaluating the robot behaviour. Often such work reveals just how little we really know, but the results are nevertheless generating relevant insights into biology, not just robot engineering.

Several groups have been pursuing this approach for some time, such as Franceschini and colleagues¹⁻⁴ in their investigation of fly visual navigation, and Cruse and associates^{5,6} studying gait control in stick insects. Robot modelling has been used to explore the function of identified neural elements within the overall context of required behavioural output. For example, work on the 'Sahabot'^{7,8} has copied the navigation capabilities of the Saharan desert ant *Cataglyphis*⁹ using polarized light detectors based on the properties of polarization-sensitive interneurons characterized by Labhart¹⁰. Similarly, the dendritic processing thought to underlie the 'looming' detection of the lobular giant movement-detector neuron in the locust¹¹ has been tested on a robot¹². A slightly different approach is illustrated in a robot model of chemotaxis of nematode worms¹³, in which the sensor and motor mechanisms of the worm were represented by robot equivalents (Fig. 1) and the neural controller tuned by an automatic adaptive process. Biological mechanisms of adaptation are themselves studied in some robot implementations, such as the investigation of the effectiveness for robot learning of mechanisms of habituation and sensitization based on the gill-withdrawal reflex in the sea slug *Aplysia*¹⁴.

Although much of the contribution of this work in clarifying and evaluating hypotheses is similar to more conventional modelling, it differs in some important



respects. Robots are required to exist within, and interact with, the real world, unlike the simplified representation of the world used in typical computer simulations. This prevents unrealistic assumptions about the available environmental signals, the nature of 'noise', or how the consequences of actions on the environment might interact with the behaviour. Robots can potentially be tested in the same natural or experimental situation as the animal itself — for example, the 'robolobster'^{15,16} was built to do chemotaxis in the same flow tank as real lobsters, producing a better understanding of the nature of the complex chemical plume they track. Robots may also be able to represent more realistically the physics of sensors and actuators by using physical copies of them. Researchers in insect walking have built robots with the same body plan as cockroaches^{17,18}, and the antennae sensors of a real moth have been used as input devices to a robot¹⁹.

A computer simulation can sometimes include more realistic detail or represent lower-level mechanisms than a robot. Nevertheless, it is becoming more common within biology to complement simulation models with physically embodied ones, so obtaining different perspectives on a problem. A more wide-ranging review of the approach can

be found in a recent review of the field of bio-inspired robotics.

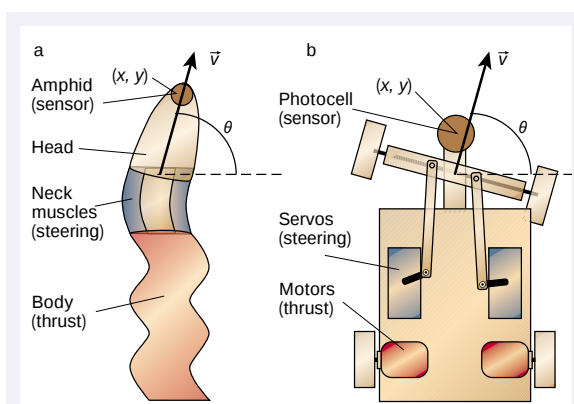


Figure 1 A robot model of chemotaxis of a nematode worm. Schematic representations of **a**, the nematode, and **b**, the robot, showing how sensory and motor systems correspond. Signal intensity is detected at a point x, y ; the direction of movement θ is controlled by head angle and the speed $\vec{v}$ by body thrust. Reprinted with permission from ref. 13.

be found in refs 20 and 21. Here I focus on work in several areas where the neural circuits are as well explored as any in neuroscience, to show some of the insights derived from the holistic, embodied view that robots provide, and how they can produce predictions for neurophysiology. I will also discuss why invertebrate neuroethology seems particularly well suited for the robot modelling approach, providing an opportunity for real progress on issues that will translate to the wider context of behavioural neuroscience.

Robot phonotaxis

Many mobile robots have been built with generic 'taxis' capabilities to approach sensory sources, often inspired by the 'thought experiments' described by Braitenberg²². Taxis behaviours have also been well studied by neuroethologists, and there is much specific data available about particular animal responses to particular signals, and the neural pathways involved. Consequently this would seem a promising area for robot modelling. Cricket phonotaxis is one example where the behaviour, neuroanatomy and neurophysiology have been studied for many years^{23–28}. Robotic implementations of this system have been investigated over the past decade^{29–32}, and improved copies of the sensory and neural processing mechanisms³³ now allow us to draw direct comparisons to the cricket.

Female crickets can locate a mate by orienting towards the species-specific calling song produced by the male. The cricket's ears are connected by a tracheal tube and thus function as pressure-difference receivers. That is, the vibration of the ear drums is the sum of direct and delayed inputs and hence is dependent on relative phase, which varies with sound-source direction for a given wavelength³⁴. The resulting difference in vibration amplitude between the ears is neurally encoded both in spike rate and spike onset latency. The characteristic temporal pattern of the sound is preserved in the spike pattern of auditory neurons and interneurons^{25,35}. The question is how the subsequent neural processing can filter the pattern to recognize the song and compare the difference between the ears to determine the direction of the singer.

The main hypothesis tested on the robot so far is that the tasks of recognition and localization may be closely interlinked. For example, using an electronic circuit to model the tracheal delays in the cricket auditory system, it was shown that a selective approach to particular carrier frequencies can result from tuning the delays to give maximal directionality for particular wavelengths of sound³¹. Use of a spiking neural network, which mimicked the temporal coding properties of identified auditory interneurons, showed that a circuit reacting to the relative latency of activation on each side would respond only to particular temporal patterns, in a way that resembled the female cricket's preference³³.

Using a robot demonstrated that this explanation of phonotaxis is viable for a real sound source in a noisy environment. It was also relatively straightforward to test the model with multiple sound sources, repeating experiments carried out on the cricket to look at the choice between similar but not identical songs, the behaviour when a song was played directly above the robot, and the reaction to a song split between two directions. The robot results were surprisingly similar to the cricket in the first two cases, suggesting that no additional recognition or decision mechanism is needed to explain these behaviours. But the results differed in the case of split song, highlighting some limitations of the model and directions for further work.

The cricket robot is intended as a test-bed for evaluating hypotheses, rather than as an engineering project with a fixed target. For example, a well-characterized pair of mutually inhibitory auditory interneurons in the cricket was not included in the model described above. How these connections might change the functionality of the system is the subject of current investigation focusing on the problems of more realistic sound fields. This means testing the robot outdoors, dealing with a large range of signal amplitudes, background noise and signal distortion from reverberation. It would be difficult to build adequate simulation models of

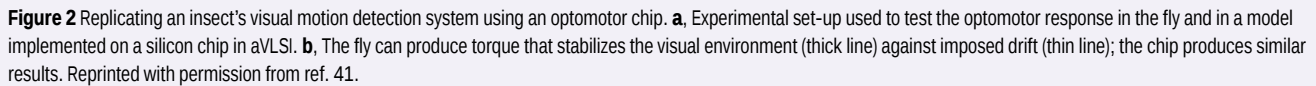
such environmental factors, but they can be easily replicated in the real world. Phonotaxis outdoors also raises the problem of how the auditory response can be integrated with other sensorimotor responses needed for movement through a complex environment. Some work in this direction is discussed below.

Robot optomotor reflex

One of the best studied areas in invertebrate neuroscience is the sensory system underlying visual motion perception in insects, so it is not surprising that this is also an area in which a number of robot models have been built. Several authors^{3,36,37} have speculated on how some rather simple but clever algorithms apparently used by insects could be adopted for controlling mobile robots, and a number of systems have been implemented. Additionally, some new discoveries in biology have been made as a result of the task-oriented perspective enforced by robotics. For example, an insect (or robot) can use motion parallax as an efficient way to avoid collisions in cluttered environments². However, such a means for detecting obstacles has the drawback that the range of effective vision decreases as the visual axis approaches the line of travel (that is, obstacles directly ahead are the hardest to see). One solution is to oscillate the direction of motion (producing a zig-zag path) or make scanning movements of the eye itself. Lewis³⁸ reports examples of zig-zag paths in insects and uses this behaviour successfully on a robot to navigate through a field of obstacles. And in a detailed investigation of the compound eye, motivated by the results of robot modelling, Franceschini and Chagneuz report³⁹ a muscle and tendon system behind the fly's eye that could produce the required scanning movement. A microscale sensor has been built based on this principle⁴⁰.

A recent study with a number of interesting features is the implementation of an optomotor aVLSI (analog very-large-scale integration) chip by Harrison and Koch^{41,42}. They use a new technology to build a physical replica of the insect's visual motion detection system, and test it by direct substitution for the insect in an experimental apparatus, and by controlling a mobile robot in an everyday environment. The technology, pioneered by Carver Mead⁴³, uses transistor circuits in the sub-threshold domain to do highly efficient, specialized calculations that exploit the inherent exponential dynamics of the silicon substrate. Harrison's circuit implements the Hassenstein-Reichardt⁴⁴ model of elementary motion detection. Each detector correlates the response at one photoreceptor with the delayed response from a neighbouring receptor. This is implemented on the chip by a multiplier operation on signals delayed by inherent lags in temporal low-pass filters. By subtracting the outputs of detectors for opposite directions, a strong direction-selective response is produced. Summing the response across all the detectors provides a signal representing full-field motion, which resembles the response of certain tangential cells in the lobular plate of the fly brain. This output is a measure of self-rotation, and is used by the fly to produce compensatory torque responses to correct deviations in heading direction — the well-known optomotor reflex.

The advantage of reproducing this model in hardware is that it performs the computationally demanding task in real time, with very low power consumption, using parallel analog computation to reduce the large bandwidth of visual input to a single, meaningful output. Hence it is possible to test the response of this system with stimulation identical to that used for experiments on the fly. The fly's optomotor response is often tested in a closed-loop flight simulator, in which the torque movements of a tethered fly in response to visual motion are fed back to control the motion of the visual scene. In this situation, flies can compensate for an imposed drift in the direction of motion (Fig. 2). Individual trials show a characteristic oscillation in torque during this behaviour. If the fly and the torque meter are then replaced by the optomotor chip, so that it receives the equivalent visual stimulus, the output from the chip can be used as a torque signal fed back to the flight simulator. In this situation, similar compensation for the imposed rotation was produced (slightly larger drift is

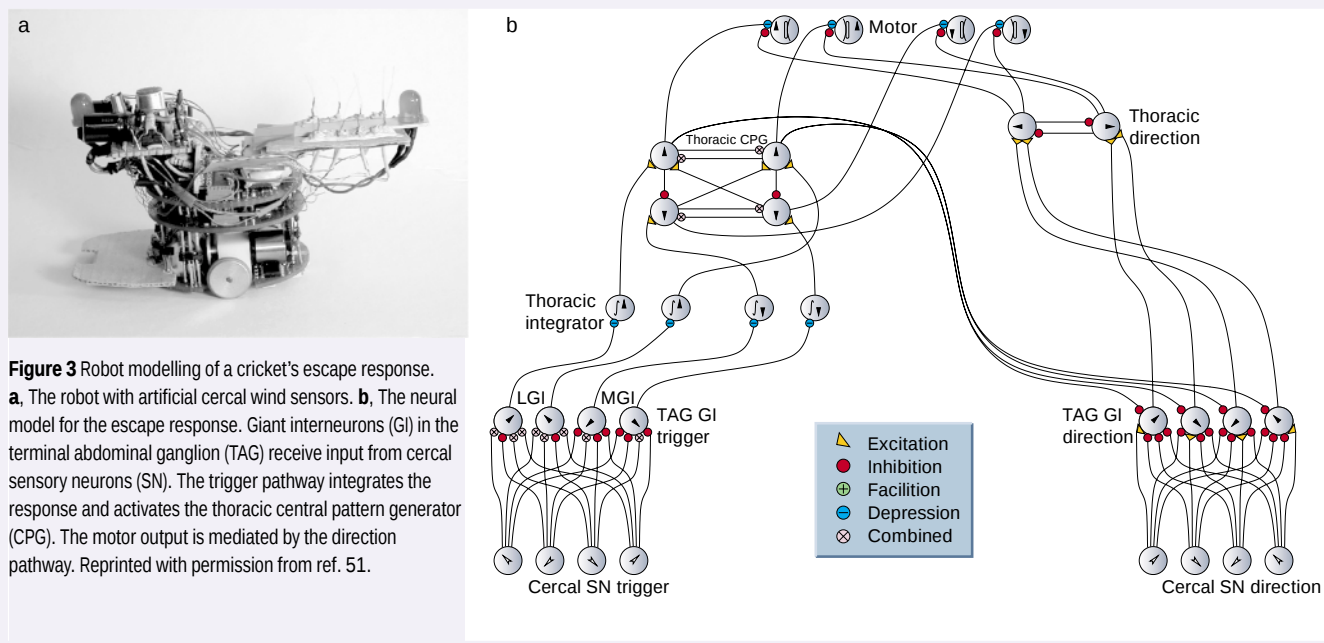


The optomotor chip is straightforward to interface to a real motor-control mechanism such as a mobile robot platform. This enables the same circuit to be tested with realistic natural input and feedback simply by driving it around in an everyday environment⁴⁵. The output of the motion processor is sufficient to instantaneously correct for as much as a 5:1 bias in left–right motor gearing, converting tight-circling behaviour without the optomotor control into straight-line motion. An unexpected bonus is that interfacing this sensory capacity to the cricket robot is also simple, allowing tests of how these behaviours might best be combined.

The output from the network model of phonotactic processing was motor commands to change the speed of the robot's left and right wheels, either making a turn towards the sound or moving forward. The optomotor chip output was added to this at the motor command stage by increasing the speed of one wheel and decreasing the other by an amount proportional to any detected visual rotation.

Two conclusions are evident from these studies. First, a precise efferent copy is impossible to produce, as it requires the system to predict the exact visual input that will occur, which is not known in a natural environment. Second, the temporal dynamics of the sensory processing and of the feedback through motor actions in the environment are critical factors to consider in evaluating the integration of different modalities. Additional constraints that may come from considering a more neurally plausible implementation of the integration mechanisms are currently being studied.

A cricket stimulated by a puff of air, such as might be created by a predator's strike, rapidly turns and runs away from the direction of the wind, producing a complete behaviour sequence in response to a short-lived stimulus. The cricket has two rear appendages (the cerci) covered in hair sensors that detect air movement. The anatomical layout and neural connectivity of the sensory axons from the cercal wind sensors has been well described^{49,50} and a small number of identified neurons well characterized. These include



four pairs of 'giant' interneurons connecting the abdominal ganglion to the motor areas of the thoracic ganglion, which are involved in initiating and steering a rapid escape. Chapman⁵¹ has built a set of direction-sensitive wind sensors that resemble the hair sensors on the cricket's cerci (Fig. 3a), and modelled the neural system using a dynamic spiking neural simulation (Fig. 3b) to produce a cricket-like escape response in a robot.

In this model, the neural pathway is divided into a 'trigger' system and a 'direction' system, which respond to 'acceleration-sensitive' and 'velocity-sensitive' hairs, respectively. The cricket has cercal hairs that vary in length, resulting in different mechanical properties, with longer hairs best deflected by a constant wind, and short hairs by a rapid puff⁵². The mechanical model hairs (essentially a fine wire attached to a spring, which when deflected in a specific direction briefly closes a contact switch) can be similarly classified. Each closure of a switch is treated as a 'spike' from a sensory neuron. This results in a low-bandwidth signal from an array of sensors that can be detected and processed at a millisecond timescale.

The neural simulation exploits the robot's microprocessor for maximum efficiency while maintaining sufficient detail to copy appropriate neuron properties. For example, synaptic delays (axon lengths) are represented by shifting a bit (a spike) along a byte. It uses a single-compartment neuron representation based on a simplified 'integrate and spike' model, and includes synaptic depression and facilitation effects. At the level of the 'thoracic ganglion' (Fig. 3b) the trigger system integrates sensory input and starts a central pattern generator circuit for forwards or backwards movement. The direction system modulates this response by inhibiting one side or the other, causing a turn. The combined output produces different kinds of turn-and-run sequences, depending on the direction of the wind stimulus, that closely resemble the cricket's behaviour⁵³. The complete network (not shown) includes input from 'antennae' (close-range distance sensors), light and sound sensors, enabling the robot to integrate these other modalities with its escape response. Thus it can avoid obstacles and follow walls while escaping, show heightened sensitivity for wind-evoked escape when light or noise levels are high, and produce an escape response to sufficiently strong changes in any one of these additional sensory cues.

This robot represents the most complete model of this neural system to date. It combines in a single coherent circuit many aspects of escape behaviour that have previously been simulated separately^{49,54}, and places the circuit within the sensorimotor–environment

loop, generating the complete behavioural response. By combining physical, behavioural and physiological constraints, particular solutions to the possible pathways become more plausible than others. Sometimes simple solutions remove potential problems. For example, by providing the robot with a cardboard buffer to make its shape more like the body of a cricket, the need for highly reliable obstacle detection was reduced as it was less likely to get stuck when it hit something. Finally, this robot shows a more complex behavioural output than the largely reactive responses of the systems described previously, by incorporating mechanisms for context-dependent plasticity of the response.

Conclusion

Robot implementations have become an accepted method for exploring issues in invertebrate neuroscience. I have described a few specific examples in detail, to show how the idea of physical modelling can contribute interesting new insights into neurobiological systems. Many other invertebrate systems have been explored in this way, and similar work is being done on a variety of vertebrate systems, including snakes, rats, fish and humans. But invertebrate systems have been a particularly successful area for the approach. Invertebrate behaviours tend to be more stereotyped and thus easier to analyse comprehensively. The number of neural connections between sensing and action is orders of magnitude less than for vertebrates, making the possibility of complete pathway mapping plausible. We should have comparable processing power available in modern computers to that in insect brains, so failure to replicate their behavioural capabilities will indicate areas in which we lack knowledge of how the systems work.

In addition, although the importance of 'embodiment' is still debated for higher cognitive processes, there is little doubt that the efficient functioning of invertebrates is highly dependent on the nature of the physical interface between their neural control systems and their environment. This is an area where robotic models have particular advantages. In a box-and-arrow, mathematical or computer model, the included constraints are only those we can think of in advance. Robot implementations introduce us to constraints inherent to the sensorimotor problem that we might otherwise fail to consider. □

1. Pichon, J.-M., Blanes, C. & Franceschini, N. in *Mobile Robots IV* (eds Wolfe, W. J. & Chun, W. H.) 44–53 (Society of Photo-optical Instrumentation Engineers, Bellingham, 1989).
2. Franceschini, N., Pichon, J. M. & Blanes, C. From insect vision to robot vision. *Phil. Trans. R. Soc.*

- Lond. B **337**, 283–294 (1992).
3. Franceschini, N. Engineering applications of small brains. *FED J.* **7**, 38–52 (1996).
4. Mura, F. & Franceschini, N. in *Intelligent Vehicles II* (eds Aoki, M. & Masaki, I.) 47–52 (MIT Press, Cambridge, MA, 1996).
5. Cruse, H., Bartling, C., Cymbalyuk, G., Dean, J. & Dreifert, M. A modular artificial neural network for controlling a six-legged walking system. *Biol. Cybern.* **72**, 421–430 (1995).
6. Cruse, H., Kindermann, T., Schumm, M., Dean, J. & Schmitz, J. Walknet—a biologically inspired network to control six-legged walking. *Neural Networks* **11**, 1435–1447 (1998).
7. Lambrinos, D. *et al.* An autonomous agent navigating with a polarized light compass. *Adapt. Behav.* **6**, 175–206 (1997).
8. Lambrinos, D., Moller, R., Labhart, T., Pfeifer, R. & Wehner, R. A mobile robot employing insect strategies for navigation. *Robot. Auton. Syst.* **30**, 39–64 (2000).
9. Wehner, R. in *Neural Basis of Behavioural Adaptations* (eds Schildberger, K. & Elsner, N.) 103–143 (Fischer, Stuttgart, 1994).
10. Labhart, T. Polarization-opponent interneurons in the insect visual system. *Nature* **331**, 435–437 (1988).
11. Rind, F. C. & Simmons, P. J. Signalling of object approach by the DCMD neuron of the locust. *J. Neurophysiol.* **77**, 1029–1033 (1997).
12. Blanchard, M., Rind, F. C. & Verschure, P. F. M. J. Collision avoidance using a model of the locust LGMD neuron. *Robot. Auton. Syst.* **30**, 17–38 (2000).
13. Morse, T. M., Ferree, T. C. & Lockery, S. R. Robust spatial navigation in a robot inspired by chemotaxis in *Caenorhabditis elegans*. *Adapt. Behav.* **6**, 393–410 (1998).
14. Dampier, R. L., French, R. L. B. & Scutt, T. W. ARBIB: an autonomous robot based on inspirations from biology. *Robot. Auton. Syst.* **31**, 247–274 (2000).
15. Grasso, F., Consi, T., Mountain, D. & Atema, J. in *From Animals to Animals 4: Proc. Sixth Int. Conf. Simul. Adapt. Behav.* (eds Maes, P., Mataric, M. J., Meyer, J. A., Pollack, J. & Wilson, S. W.) 104–112 (MIT Press, Cambridge, MA, 1996).
16. Grasso, F., Consi, T., Mountain, D. & Atema, J. Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: progress and challenges. *Robot. Auton. Syst.* **30**, 115–131 (2000).
17. Quinn, R. D. & Ritzmann, R. E. Construction of a hexapod robot with cockroach kinematics benefits both robotics and biology. *Connect. Sci.* **10**, 239–254 (1998).
18. Delcomyn, F. & Nelson, M. E. Architectures for a biomimetic hexapod robot. *Robot. Auton. Syst.* **30**, 5–15 (2000).
19. Kuwana, Y., Shimoyama, I. & Miura, H. in *Proc. IEEE Int. Conf. Intell. Robots Syst.* 530–535 (IEEE Computer Society Press, Los Alamitos, CA, 1995).
20. Webb, B. What does robotics offer animal behaviour? *Anim. Behav.* **60**, 545–558 (2000).
21. Webb, B. Are 'biorobots' good models of biological behaviour? *Behav. Brain Sci.* (in the press).
22. Braitenberg, V. *Vehicles: Experiments in Synthetic Psychology* (MIT Press, Cambridge, MA, 1984).
23. Huber, F. Behavior and neurobiology of acoustically oriented insects. *Naturwissenschaften* **79**, 393–406 (1992).
24. Huber, F. & Thorson, J. Cricket auditory communication. *Sci. Am.* **253**, 47–54 (1985).
25. Wohlers, D. W. & Huber, F. Processing of sound signals by six types of neurons in the prothoracic ganglion of the cricket *Gryllus campestris* L. *J. Comp. Physiol.* **146**, 161–173 (1981).
26. Schildberger, K. Behavioural and neuronal methods of cricket phonotaxis. *Experientia* **44**, 408–415 (1988).
27. Wendler, G. in *Sensory Systems and Communication in Arthropods* (eds Gribakin, F. G., Wiese, K. & Popov, A. V.) 387–394 (Birkhäuser, Basel, 1990).
28. Pollack, G. S. in *Comparative Hearing: Insects* (eds Hoy, R. R., Popper, A. N. & Fay, R. R.) 139–196 (Springer, New York, 1998).
29. Webb, B. in *From Animals to Animals 3: Proc. Third Int. Conf. Simul. Adapt. Behav.* (eds Cliff, D., Husbands, P., Meyer, J.-A. & Wilson, S. W.) 45–54 (MIT Press, Cambridge, MA, 1994).
30. Webb, B. Using robots to model animals: a cricket test. *Robot. Auton. Syst.* **16**, 117–134 (1995).
31. Lund, H. H., Webb, B. & Hallam, J. in *Fourth Eur. Conf. Artif. Life* (eds Husbands, P. & Harvey, I.) 246–255 (MIT Press, Cambridge, MA, 1997).
32. Lund, H. H., Webb, B. & Hallam, J. Physical and temporal scaling considerations in a robot model of cricket calling song preference. *Artif. Life* **4**, 95–107 (1998).
33. Webb, B. & Scutt, T. A simple latency dependent spiking neuron model of cricket phonotaxis. *Biol. Cybern.* **82**, 247–269 (2000).
34. Michelsen, A., Popov, A. V. & Lewis, B. Physics of directional hearing in the cricket *Gryllus bimaculatus*. *J. Comp. Physiol. A* **175**, 153–164 (1994).
35. Schildberger, K. Temporal selectivity of identified auditory interneurons in the cricket brain. *J. Comp. Physiol.* **155**, 171–185 (1984).
36. Horridge, G. A., Loughet-Higgins, H. C. & Horridge, G. A. What can engineers learn from insect vision? *Phil. Trans. R. Soc. Lond. B* **337**, 271–282 (1992).
37. Srinivasan, M. V. & Venkatesh, S. *From Living Eyes to Seeing Machines* (Oxford Univ. Press, Oxford, 1997).
38. Lewis, A. in *Advances in Neural Information Processing Systems 10* (eds Jordan, M. I., Kearns, M. J. & Solla, S. A.) 822–828 (MIT Press, Cambridge, MA, 1998).
39. Franceschini, N. & Chagneux, R. in *Neurobiology: From Membrane to Mind. Proc. 25th Gottingen Neurobiol. Conf.* (eds Elsner, N. & Wasse, H.) (G. Thieme, Stuttgart, 1997).
40. Hoshino, K., Mura, F. & Shimoyama, I. Design and performance of a micro-sized biomorphic compound eye with a scanning retina. *J. Microelectromech. Syst.* **9**, 32–37 (2000).
41. Harrison, R. R. & Koch, C. A silicon implementation of the fly's optomotor control system. *Neural Comput.* **12**, 2291–2304 (2000).
42. Harrison, R. R. & Koch, C. in *Advances in Neural Information Processing Systems 10* (eds Jordan, M. I., Kearns, M. J. & Solla, S. A.) 880–886 (MIT Press, Cambridge, MA, 1998).
43. Mead, C. *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, MA, 1989).
44. Hassenstein, B. & Reichardt, W. Systemtheoretische Analyse der Zeit-, Reihenfolgen-, und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers. *Chlorophanus. Z. Naturforsch.* **11b**, 513–524 (1956).
45. Harrison, R. R. & Koch, C. A robust analog VLSI motion sensor based on the visual system of the fly. *Auton. Robot.* **7**, 211–224 (1999).
46. Weber, T., Thorson, J. & Huber, F. Auditory behaviour of the cricket I. Dynamics of compensated walking and discrimination paradigms on the Kramer treadmill. *J. Comp. Physiol. A* **141**, 215–232 (1981).
47. Böhm, H., Schildberger, K. & Huber, F. Visual and acoustic course control in the cricket *Gryllus bimaculatus*. *J. Exp. Biol.* **159**, 235–248 (1991).
48. Webb, B. & Harrison, R. R. in *Proc. SPIE Symp. Sensor Fusion and Decentralized Control in Robotic Systems III* (eds McKee, G. T. & Schenker, P. S.) 113–124 (SPIE, Boston, MA, 2000).
49. Kohstall-Schnell, D. & Gras, H. Activity of giant interneurons and other wind sensitive elements of the terminal abdominal ganglion in the walking cricket. *J. Exp. Biol.* **193**, 157–181 (1994).
50. Paydar, S., Doan, C. & Jacobs, G. Neural mapping of direction and frequency in the cricket cercal system. *J. Neurosci.* **19**, 1771–1781 (1999).
51. Chapman, T. *Morphological and Neural Modelling of the Orthopteran Escape Response*. Thesis, Univ. Stirling (2001).
52. Shimozawa, T. & Kanou, M. The aerodynamics and sensory physiology of range fractionation in the cercal filiform hair of the cricket *Gryllus bimaculatus*. *J. Comp. Physiol. A* **155**, 495–505 (1984).
53. Tauber, E. & Camhi, J. The wind-evoked escape behaviour of the cricket *Gryllus bimaculatus*: integration of behavioral elements. *J. Exp. Biol.* **198**, 1895–1907 (1995).
54. Ezrachi, E., Levi, R., Camhi, J. & Parnas, H. Right-left discrimination in a biologically oriented model of the cockroach escape system. *Biol. Cybern.* **81**, 89–99 (1999).

All Creatures Great and Small

NATIONAL INSTITUTES OF HEALTH



National Institute of Neurological Disorders and Stroke

Modeling the Nervous System

The mission of the National Institute of Neurological Disorders and Stroke (NINDS) is to reduce the burden of neurological disease. To achieve this goal, NINDS has supported and conducted research on the healthy and diseased nervous system for more than 50 years. A critical component of this research program has been the use of non-mammalian model organisms for investigating neural development and function. NINDS-supported investigators routinely use yeast, flies, worms, birds, crabs, fish and a variety of other animals in these studies.

NINDS-funded researchers have used non-mammalian models to make important breakthroughs in many areas of neuroscience. This is particularly true in the field of neural development. *C. elegans*, *Drosophila*, zebrafish, and other animals have been essential tools for dissecting the molecular events that control neurogenesis, axon guidance, synapse formation, and other developmental processes. Understanding these events is important not only from a basic scientific perspective, but because it provides a rational basis for developing treatments for neurological disease. Studies of neurogenesis, for example, are illuminating the therapeutic potential of human stem cells. Identification of the molecules that guide early axon outgrowth provides insight into how damaged connections can be regenerated or repaired. Increased understanding of cell division facilitates the development of treatments for brain tumors, neurofibromatosis, and other diseases characterized by unrestricted cell proliferation.

As described in this *Nature* Insight, model organisms are also critical for understanding the neural circuitry and synaptic physiology that underlie brain functions. Investigators are modeling learning, circadian rhythms, sleep, pain, and other higher-order properties in lower species. They are also using these organisms to identify and study the channels, neurotransmitters, and other molecules that ultimately control these neural systems. By studying how such functions are controlled in simpler nervous systems, we are obtaining clues about what goes wrong in epilepsy, autism, ADHD and other brain disorders.

An area where non-mammalian models hold tremendous promise is in developing treatments for neurodegenerative disease. Parkinson's disease, stroke, Alzheimer's disease, Huntington's disease, amyotrophic lateral sclerosis (ALS), spinal muscular atrophy, muscular dystrophies and many other disorders cause cell death in specific regions of the nervous system. NINDS-supported researchers have created genetic models of many of these disorders using worms, flies, fish, and other organisms. Such models are being used to understand what causes cell death and, more recently, to screen candidate therapeutics.

This is a time of tremendous progress and increasing hope in the battle against neurological disease. Greater understanding of the nervous system is beginning to pay off in the form of treatments for previously intractable problems including spinal cord injury, acute stroke, epilepsy, multiple sclerosis, Parkinson's disease, and other disorders. Non-mammalian models will continue to be invaluable in this process.

For more information, contact Robert Finkelstein, Ph.D., +1 301-496-5745, finkelsr@ninds.nih.gov.



National Institute of Child Health and Human Development

Non-Mammalian Model Organisms in Neural Development

NICHD's mission is to assure that every individual is born healthy and wanted. Current areas of biomedical and behavioral research interest include normal and abnormal brain development, function and behavior, neural tube formation and defects, repair and recovery of motor and cognitive function, and nutritional and hormonal effects on brain development. The goal is to improve understanding of underlying mechanisms associated with birth and developmental defects, and through this knowledge, learn how to prevent and treat them. To achieve this aim, NICHD-supported scientists rely on non-mammalian model organisms, including fruit fly, honeybee, zebrafish, frog, chick, and songbird, as investigative tools.

The zebrafish, *Danio rerio*, is an important genetic model of vertebrate development. We now know that gene expression patterns define developmentally distinct domains in axial structures and in the nervous system. Mutations have been identified that alter expression patterns and axial morphogenesis in these embryos, and a linkage map of the zebrafish genome, including many of these mutations and molecular markers, has been generated. Genetic and RH maps and EST projects have facilitated the cloning of mutant loci. The recent development of morpholinos and other genetic tools, the increase in novel mutants obtained from mutagenesis screens, and recent advances in *in vivo* imaging technologies will provide new insights into the genetic basis of neural development and behavior. NICHD and NIDDK co-chair the Trans-NIH Zebrafish Coordinating Committee (www.nih.gov/science/models/zebrafish), which actively promotes the zebrafish at NIH and in the scientific community.

The *Xenopus* embryo has long served as a major model for the study of embryonic development. Using *Xenopus*, NICHD-supported investigators have a detailed cellular and molecular understanding of early developmental events, including patterning of the basic body plan, the determination of cell fate, and the early patterning of major organ systems, including the nervous system.

continued on page 364

In addition, many of the factors originally identified in *Xenopus* have been subsequently shown to control later developmental events, as well as other critical biological processes. NICHD also chairs the Trans-NIH *Xenopus* Committee, which has supported the production of genetic and genomic resources including a collection of cDNA libraries, an EST database, a UniGene Set, and several BAC libraries. Additionally, projects to mutagenize and phenotype *Xenopus tropicalis* have just been initiated. Information about these resources and about future plans is available at www.nih.gov/science/models/xenopus. The fruit fly, *Drosophila*, has also classically served to increase our understanding of pattern formation, segmentation, and cellular migration in the developing nervous system. Recent work has investigated a gene and protein in the fruit fly that share many of the gene and protein alterations found in Fragile X syndrome, the most common inherited form of mental retardation. Modification of the gene in the fly leads to morphological abnormalities in neuronal processes, a finding similar to that seen in the CNS of individuals with Fragile X syndrome. These changes are accompanied by altered neurotransmission. This research provides insight into the developmental sequence of events that may underlie the genesis of altered neuron function in individuals with Fragile X syndrome.

Avian models such as the songbird have been studied to understand how brain and behavior are linked, as well as to identify what factors act on the brain to regulate the development of behavior. NICHD-supported researchers using avian models have investigated the impact of hormones on developing brain organization and behaviors such as courtship, mate choice, aggression, and spatial memory. Researchers are investigating the influence of testosterone on certain sexually selected traits in male birds. Variation in these traits is associated with the quality of parenting in male birds. In avian models, hippocampal formation is being intensely investigated because of its critical involvement in learning and memory in both birds and mammals.

For more information, contact Deborah Henken, Ph.D., +1 301-435-6885, dh50g@nih.gov.

NIDA National Institute on Drug Abuse *Non-Mammalian Models in Drug Abuse Research*

NIDA supports 85 percent of the world's research on drug abuse and addiction. Drug addiction is a relapsing disorder characterized by loss of control over drug intake and a persistent desire to use a drug. Relapse can occur after long periods of abstinence and may result from stress and/or re-exposure to a drug or to environments associated with drug use. These characteristics, together with a large body of research, indicate that repeated exposure to addictive drugs produces long-lasting cellular and molecular brain changes.

Studies of human addicts and in animals have identified key neural circuits, intracellular pathways, and environmental and genetic factors involved in drug abuse and addiction. Addictive drugs are neurochemically diverse and affect numerous receptors and cellular processes, but they share the ability to activate the mesolimbic dopamine system, either directly or indirectly. Similar to natural reinforcers such as food and sex, drug activation of this system mediates their initial rewarding properties and the associations formed between drugs and environmental cues. Unlike natural reinforcers, however, repeated exposure to drugs can produce abnormal neuroadaptations that may be responsible for the unusual strength and duration of their ability to alter motivated behavior. In addition to their shared effects, each drug can produce unique, widespread, and complex changes in the CNS through chronic activation of specific neurotransmitter receptors and cellular processes. Thus, addictive drugs can affect the CNS via normal learning mechanisms, by provoking homeostatic compensation in response to over- or underactivation of receptors and signal transduction pathways, and by neurotoxic effects, impairment of neurogenesis, or disruption of plasticity and repair.

Although mammalian models are becoming increasingly sophisticated in their ability to explain human drug use, the recent identification of multiple cellular responses to drugs, such as alterations in gene expression, upregulation of cAMP pathways, and long-lasting structural changes in dendrites, has expanded the value of non-mammalian models in drug abuse research. Studies in species with simple or specialized nervous systems or behaviors, or in genetic model organisms, are relevant to NIDA's mission.

Animal models of learning, memory, and motivated behaviors can provide insight into the neurobiology of addiction and the effects of drugs on these processes, regardless of species. For example, NIDA-sponsored research has shown that song learning is impaired in zebra finches exposed to cannabinoids and that zebrafish prefer environments associated with cocaine. Well-characterized non-mammalian nervous systems are also valuable for investigating the mechanisms of drug-induced changes in dendritic architecture, or to model the results of altered neuromodulation in circuits for fixed action patterns.

Forward genetics is a powerful approach to delineate biological processes underlying drug abuse and addiction. The short generation times and well-characterized genetics of *Drosophila*, *C. elegans*, and zebrafish make them ideal for producing mutations affecting behavioral phenotypes and identifying genes. Genetic screens can rapidly identify genes and signal transduction pathways that regulate acute and chronic effects of drugs such as nicotine, cocaine, amphetamines, cannabinoids, opiates and inhalants. For example, NIDA investigators have shown that phosphorylation of nicotinic receptors mediates adaptation to nicotine in *C. elegans*, and that gene products regulating circadian rhythms are necessary for sensitization to cocaine in *Drosophila*.

For further information about NIDA's interest in non-mammalian neurobehavioral systems, contact Susan Volman, Ph.D., +1 301-435-1315, svolman@nida.nih.gov. For information about NIDA's interest in genetic organisms, contact Jonathan D. Pollock, Ph.D., +1 301-435-1309, j183r@nih.gov.



National Institute on Deafness and Other Communication Disorders

Modeling Hearing, Language, Smell and Taste

NIDCD-supported research has traditionally included many non-mammalian models for the study of hearing, balance, voice, speech, language, smell and taste. Many investigators take advantage of the less complex auditory systems of non-mammalian organisms (insects, birds, anurans, fish and spiders) to study the neurobehavioral mechanisms that mediate the recognition, communication, and localization of acoustic signals. Similar to mammals, these model organisms use acoustic signals to mediate behaviors such as mate recognition, courtship, territoriality, and detection and evasion of predators. Findings from these studies provide model systems to understand the evolution of these crucial behaviors, facilitating their application to human health.

One example of how basic research in non-mammalian models can translate into applied research in future generation hearing aids is the development of a directionally sensitive microphone based on the structure of a parasitic fly ear. In this fly, *Orima ochracea*, the ears are so close together that interaural time and intensity differences are minute, yet these flies localize sounds quite well. These insect ears have been mimicked through microelectromechanical systems to construct micro-scale, passive mechanical structures for improved nano/micro-scale directional microphones in hearing aids. In related studies, male frogs produce advertisement calls in large choruses and females must identify and localize the individual callers based on spectrotemporal characteristics of sounds in a noisy environment. This work will help determine the mechanisms by which the sounds of individual callers are segregated in the central auditory system. Such findings will guide the design of "intelligent" hearing aids able to solve the significant problem of listening in noise.

Birdsong is a model that parallels human speech learning. Song is an intricate motor act that is learned by young birds in two phases: the bird memorizes a tutor song (sensory learning); the bird then begins to sing, and, like human

infants, uses auditory feedback to refine its vocalizations until they match the memorized tutor song (sensorimotor learning). These features give birdsong the potential to shed light on the neural basis of learning, and on factors that control and limit human learning.

Research on taste and smell has historically involved a variety of non-mammalian model systems. These phylogenetically older sensory systems are physically larger, more accessible, and possess a less complex organization than their mammalian counterparts. Research has involved insects (moth, honeybee, mosquito and fly), roundworms, fish (zebrafish, goldfish, carp and catfish), amphibians (frog and salamander), and reptiles (turtle and snake). Approaches including electrophysiology, behavior, functional imaging, morphology, molecular biology and genetics have been used to study the function, structure and genomic features of the chemical senses. For example, recent studies in the fruit fly have identified the separate families of taste and smell olfactory receptor genes, which encode G-protein-coupled receptors that respond to odors and tastants, respectively.

For more information, contact Lynn E. Luethke, Ph.D., +1 301-402-3458, lynn_luethke@nih.gov.



National Institute of General Medical Sciences

Model Perspectives of the Brain and Behavior

The mission of the National Institute of General Medical Sciences (NIGMS) is to support research and research training in the basic biomedical sciences. Almost 90 percent of the NIGMS budget for these programs is directed to the support of investigator-initiated research in the most basic and fundamental areas of biomedical science. Such research provides the foundation for subsequent disease-targeted studies supported by the other components of the NIH. Much of this fundamental research involves the use of model systems such as fruit flies (*Drosophila*), roundworms (*C. elegans*), zebrafish and "simple" organisms like yeast (*S. cerevisiae*) and fungi (*Neurospora*) to focus on molecular aspects of cellular physiology, genetic regulation and developmental processes. Each year, NIGMS-supported scientists make major advances in understanding fundamental life processes. In the course of answering basic research questions, these investigators also increase our knowledge about the mechanisms involved in certain diseases. Other grantees develop important new tools and techniques, many of which have applications in the biotechnology industry. In recognition of the significance of their work, a number of NIGMS grantees have received the Nobel Prize and other high scientific honors.

Research in neuroscience supported by NIGMS is broad in scope. It primarily involves the use of non-mammalian model systems to examine the genetic regulation and biochemical mechanisms underlying processes such as biological rhythms. Model systems enable geneticists and neurobiologists to identify and characterize genes that are involved in all aspects of the developing, mature and aging nervous system by characterizing the behavior, physiology and biochemistry of mutant animals.

An excellent example of the power of non-mammalian model systems is illustrated by the seminal advances made by NIGMS-supported investigators and others in discovering the basic blueprint for the circadian clock using fruit flies, fungi (*Neurospora*) and plants (*Arabidopsis*). Establishment and regulation of circadian rhythms is under the control of several genes that we now know are highly conserved in evolution. The core circadian oscillator in fungi, flies, mice and humans is composed of a transcription/translation negative feedback loop. In this system, the rhythmic expression of clock genes results in cycling levels of inhibitory RNAs and proteins. This detailed understanding of the biological clock will eventually lead to therapeutic approaches to treating circadian related disorders. The same model systems that have already provided critical insights will continue to advance basic and applied research in this area.

In addition to a focus on circadian biology, NIGMS-supported investigators are using *Drosophila*, *C. elegans* and other models to identify and characterize genes and gene interactions that are involved in neural development, sensory transduction, synaptic plasticity, learning and memory, reproductive behavior and aggression. NIGMS-supported work on the regulation of neurotransmitter synthesis, transport and receptors has advanced our understanding of physiological changes that may underlie drug addiction.

Two program announcements that are components of the NIGMS Initiative on Studies of Complex Biological Systems are relevant to the application of model systems in neuroscience: Quantitative Approaches to the Analysis of Complex Biological Systems, and Genetic Architecture of Complex Phenotypes. With these program announcements, and through investigator-initiated projects, NIGMS will continue its support of research that explores fundamental aspects of nervous system function using model systems.

For more information, contact Michael Sesma, Ph.D., +1 301-594-2048, ms265g@nih.gov.



National Institute of Mental Health

Modeling Neurobehavioral Disorders

The National Institute of Mental Health (NIMH), as part of its mission to improve clinical diagnosis and develop better strategies for the treatment and prevention of mental disorders, supports basic neuroscience research. The past few years have witnessed extraordinary advances in our understanding of nervous system functions and complex behaviors at the genetic, molecular and cellular level. Much of this new knowledge and insights have come from research conducted in various non-mammalian model organisms, including *C. elegans*, *Drosophila*, *Xenopus*, zebrafish and zebrafinch. Studies in model systems have already provided important insights into our understanding of mechanisms, pathways and molecules involved in normal brain function and in a variety of human diseases. A number of recent advances in developmental neuroscience have been achieved as the direct result of critical studies in non-mammalian systems, delineating transcription factor networks, neuronal-glial interactions and axon guidance mechanisms that underlie the establishment of neural circuitry and behavior in mammalian systems. Genetic studies in organisms such as yeast and *C. elegans* have identified novel genes, signaling molecules, and molecular mechanisms of action of drugs used in the treatment of mental disorders. NIMH-supported research studies in *Drosophila* have led to exciting discoveries of critical importance on molecular components of biological clocks, detailed descriptions of how different regulatory networks function, and the cellular basis of rhythmicity. Another important area of interest to NIMH is the determination of gene networks, cellular pathways and neural circuits for learned behavior, for which learned vocal communication in zebrafinch represents an ideal experimental paradigm. NIMH is interested in promoting and supporting molecular, cellular, behavioral and genetic neuroscience research utilizing non-mammalian model organisms in all aspects of neural functioning and behavior, including but not limited to, circadian rhythms, sleep, functional neuroanatomy, signal transduction, neuroendocrinology, neuroimmunology, psychopharmacology, neurodevelopment, cognition, emotion, learning and memory, behavioral regulation at the cellular and system level, and the genetic basis of complex behaviors. NIMH supports and conducts research and research training in the normal and disordered neurobiological processes implicated in mental disorders. The NIMH is working closely with other NIH institutes and federal agencies to promote both basic and clinical biomedical research in laboratories across the country through a peer-reviewed program of grants and contracts. The Institute also conducts research in intramural laboratories and clinical facilities, and collaborates with other NIH institutes and federal agencies.

For further information, contact Hemin R. Chin, Ph.D., +1 301-443-1706, hchin@mail.nih.gov.

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

Production Manager: Billie Franklin

To send materials use London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@nature.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager: Kate Cowan

naturejobs

A basic problem

It is no longer news that the number of scientists trained to translate basic research into clinical practice has been declining over the past two decades. And that, at the same time, the huge expansion in genomic and proteomic data has meant that the quantity of things to translate has increased enormously. The career problem, crudely stated, is there are too many basic biologists and not enough physician scientists.

Participants at a 'Days of Molecular Medicine' conference this March agreed that one problem might be that the traditional MD/PhD pathway is no longer the best way to train scientists who want to translate findings from the bench to the bedside. Such programmes take too long, cost too much, and offer too few slots to meet the existing need.

The answer, the conference participants suggested, is to find different routes and incentives. *Nature Medicine*, which organized the meeting with the University of California, San Diego, and the Salk Institute, this month published three solutions that the conference came up with (full text is on the *Naturejobs* Life Sciences Channel).

One is to close the gap from the 'other side', by giving some PhDs more medical training, even allowing them to spend time on rounds with medical residents (see *Nature Med.* **8**, 433–436; 2002). Another solution is to find ways to encourage more 'late bloomers' — medical students who decide midway through their training that they have an interest in research (see *Nature Med.* **8**, 437–439; 2002). Finally, as women are disproportionately under-represented in the MD/PhD ranks, policy-makers should look at ways to reverse this trend (see *Nature Med.* **8**, 439–441; 2002).

Although no one approach is likely to be a panacea, combining several might help biologists to find jobs and turn data into drugs.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Expansion and investment help Boston stay attractive p4

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



A magnetic hub Boston

The Boston area is now a major region for biotechnology research.

Scientists working in Boston and Cambridge, Massachusetts, sometimes refer to the area as 'the hub'. But 'the magnet' might be a more suitable name — or 'the bank'. The greater Boston area leads the United States in attracting money from the National Institutes of Health (NIH) — it received \$1.5 billion in 2000, nearly 10% of the biomedical agency's entire budget for that year.

Harvard University and the Massachusetts Institute of Technology (MIT), the area's two flagship universities and their affiliates, draw so many donations that anything under \$100 million barely registers. And the area, which already has one of the largest concentrations of biotechnology companies in the world — and a few billion dollars in venture capital funding — is now attracting the interest of more established drug companies. Novartis, for example, announced this month that it is moving its worldwide research and development headquarters to Cambridge.

All of these components add up to massive job opportunities, although the cost of living in the area is rising and the biotech market remains unstable after the terrorist attacks of 11 September. And, despite the implication that all the sites are in contact with each other in the 'hub', the area has a reputation for competition and secrecy that some researchers say is not without foundation.

Susan Lindquist, director of the Whitehead Institute in Cambridge, believes that with "spectacular growth" ahead and a reduction in competition — at least for resources — some of those boundaries may be dissolving. Lindquist is currently evaluating whether to split her genome institute into two — one part for genome sequencing, the other to do more postgenomic work, using tools such as microarrays



and mass spectrometers. She thinks that such a split might attract a wider variety of scientists and allow the institute to capitalize on the data it is generating.

Lindquist is also considering how to create more flexible career paths within the institute, possibly by balancing tenure-track positions with more short-term contracts. In addition, she wants to find scientists who will run and improve technological tools, in much the same way that physicists at large labs tend instrumentation but are still regarded as scientists rather than "high-priced technicians", she says.

Shuguang Zhang, associate director of MIT's Center for Biomedical Engineering is also breaking down barriers. For him, the interdisciplinary approach is what makes the centre run. Zhang, a geneticist by training, is trying to design biopolymer scaffolds onto which cells can be cultured to generate replacement tissues. For his work, he relies on collaborations with mathematicians, engineers, computer scientists, and even a



Susan Lindquist



Shuguang Zhang

Joining the drug hunt

The Boston area, with its combination of universities, research hospitals and venture capital, is attracting increased activity in cancer-drug discovery. About a quarter of Boston's 150 or so biotech companies engaged in drug discovery have active cancer-research programmes.

One of the newest additions, GenPath, is a good illustration of the local synergy. The company was founded by Dana-Farber Cancer Institute researchers Ronald DePinho and Lynda Chin. It is funded with \$15 million from local venture capital firm MPM Capital, which has over \$1 billion under management. And it is advised by local cancer-research luminaries, including Tyler Jacks, director of the Center for

Cancer Research at the Massachusetts Institute of Technology, and Raju Kucherlapati, director of Harvard Partners Center for Genetics and Genomics. The company expects to hire over 20 scientists in the next 6–9 months.

Meanwhile, AstraZeneca's Boston R&D facility illustrates how more established companies are increasing their activity in the area. When Jeffrey Hanke, the company's vice-president of cancer research, joined AstraZeneca about two years ago, he oversaw a group of seven staff. That number has grown to about 50, and will double once they move into a new building to be completed late next year. Infectious diseases and technology development groups will share the new facility. **P.S.**

scientist from MIT's famed media lab. To foster these links, he frequently crosses the sky bridges that connect his biological engineering building with the chemical engineering and biology buildings. He even crosses rhetorical bridges to work with George Whitesides, chemist and biopolymer expert at 'rival' Harvard.

EXPANSION PLANS

Harvard, too, is showing signs of building bridges as it expands. The university is in the second year of a \$700-million, five-year building plan (see *Nature* **409**, 271; 2001), fuelled partly by the NIH — Harvard received \$250 million in 2000.

One of its latest additions is the Bauer Center for Genomics Research, which aims to conduct postgenomics research by teaming physicists, mathematicians, chemists and computer scientists with biologists (see *Nature* **416**, 256–257; 2002). The centre, which is still searching for a few more fellows, puts its emphasis on collaboration. And, like Zhang's lab, its location is designed to stimulate those interactions. It serves as a node between a biochemistry building and a building for molecular and cellular biology.

Another recently announced project in the area is a \$50-million genetics building that will house 300 research staff shared by Harvard Medical School, Brigham and Women's Hospital and Massachusetts General Hospital. Denise Faustman, an immunologist at Massachusetts General Hospital, notes that the

concentration of institutes makes recruiting easier — especially for candidates who have scientist spouses, as their partner's employment prospects are also improved. Once graduate students or postdocs secure a position, the area's reputation gives them a head start over young scientists from some other parts of the country.

Faustman gets regular calls from recruiters wanting to employ people from her lab.

The high concentration of universities, hospitals and drug-development companies in the Boston area also provides job security and allows some to take risks earlier in their career. Faustman has hired a few scientists whose gambles in younger biotech companies have not worked out.

SECURE POSITIONS

The perception of job security helped Scott Sneddon decide to risk moving from established drug firm Pfizer in Connecticut to the smaller biotech company Genzyme in Cambridge. He counsels his friends in the drug industry to take similar chances in the area.

He says that he does not regret his move, because he enjoys his new employer's interactions with academia. For example, David Housman, an MIT researcher who specializes in Huntington's disease, approached Genzyme with an assay that he thought might be useful to screen for the disease. The company tweaked the assay and produced a more refined screen, and in the process Housman gained a better understanding of Huntington's.

Although location in the area does result in such synergies, it can also cause some difficulties, says Connie Pate, senior director of recruitment at Millennium Pharmaceuticals. As the company shifts from drug-discovery into product development, it is looking for more "downstream" people, she says — scientists with expertise in regulatory affairs, biostatistics and clinical trials.

With more drug firms and other established biotech companies moving in, finding qualified local candidates is getting tougher, says Pate. "We have to cast the net wider." That means appealing to scientists from parts of the country with significantly lower costs of living, she says.

Still, even with such pressure, Peter Blume-Jensen had little problem finding qualified candidates for his fledgling cancer-research group at the Serono Reproductive Biology Institute in Randolph. Like many other companies in the area, Serono, a specialist in fertility medicine, is now investing heavily in cancer-research drug discovery (see 'Joining the drug hunt'). Even though Blume-Jensen cast an international net, he ended up hiring several scientists already working in the Boston area, showing that the 'magnet' has already managed to attract a huge amount of talent. ■

Paul Smaglik is editor of *Naturejobs*.

Whitehead Institute ♦ www.wi.mit.edu

Center for Biomedical Engineering ♦ web.mit.edu/cbe/www

Bauer Center for Genomics Research ♦ cgr.harvard.edu



The Whitehead Institute (above) hopes to capitalize on the genomics data it is generating.